

QY 463 GAGAAGATTATGAGGATCGCAAGCCAGGTCCTTGGTGTCTCAACTGGACCATGGCC 522
|||||
Db 5669 GAGAAGTCTAAGATGACAGGATTGGCCAGTCCATCGGGGTGTCCAACTTCAACACAGG 5728
QY 523 GACCTTGAGAGATGTCCTCAAGTTGCG-----CAAGGTCACTGCTCAAGCCACCATC 576
|||||
Db 5729 CTGCTGGAGATGATCTTCAACAGCCAGGCTCAAGTACAGGCTGTCTGCAACACAGGTG 5788
QY 577 GAGATTACCCCTTCTCTCCCAACAGGAGCTGGTGCAGTACTGCTTCTCCAAAGACATT 636
|||||
Db 5789 GAAATGTCATCTTACTTCAACAGAGAAACTGCTGGATTCTGCAAGTCAAGAGACATT 5848
QY 637 ATGCCCGTGGCTACTCTCTCTGGCTGCTGAGAACACAGGTTC-----CACACC 687
|||||
Db 5849 GTTCTGGTGGCTATAGTCTCTGGATCCATCGAGAACCAACGATGGGTGACCCGAAC 5908
QY 688 GGTGAGCGGTGAGGAGAAAGACTCTGAACAGATCGCGAGAGAGGGGGCAACACC 747
|||||
Db 5909 TCCCGGTGCTCTGGAGAACAGTCTTGTGCTTGGCAAAAGACACAGCGAAC 5968
QY 748 CTGCTCAGGTTCTTATTGCTGGGTCTGGCGCTGCTGCTACGTCTGCTTCTCCCAAGAGC 807
|||||
Db 5969 CCAGCCCTGATTGCTTCAACAGAGAAACTGCTGGATTCTGCAAGTCAAAAGACATT 5848
QY 808 TCCACCCCAAGCGATTGATCCAACTTCAAGAGCATGAGCTC 852
|||||
Db 6029 TACAATGAGCGCGCATCAGACAGAACGTGAGGTGTTGAATTC 6073

RESULT 15

US-09-614-124B-784
; Sequence 784, Application US/09614124B
; Patent No. 630574
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedwick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614,124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 784
; LENGTH: 6353
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-614-124B-784

Query Match 10.9%; Score 106.6; DB 4; Length 6353;
Best Local Similarity 50.5%; Pred. No. 7e-21;
Matches 386; Conservative 0; Mismatches 349; Indels 30; Gaps 4;
QY 103 GCTGTCAACACTGCTGAGAGCCGGTTACCGTCACTTGGACTGTGCTGTACTACTG 162
Db 5324 GCGGTCAAATGGCAATGAGACCGGGTTCACCATATTGATTCTGCACATGTTTACAAT 5383
QY 163 AACGAGGTGAGGTGGTGGAGGTATCCGTGACTTCTCTGAAGAGAACCCCTCGTGAAG 222
Db 5384 AATGAGGAGCAGTTGGACTGGCCATCG---AAGCAAGATTGCAGATGGCAGTGTGAAG 5440
QY 223 CGTAGGACATCTTGGTCTGACCAAGGTGTGGAACCACTCCACCGTTATGAGGAGCTC 282
Db 5441 AGAGAAGACATATTCTACACTTCAAAGCTTTGGAGCAATTCCTCATCGACAGATTGGTC 5500
QY 283 CTCTGCTCATTTGACGACTCCCTGAAGCGTCTTGGACTGTGACTAGTGTGATGTTCTC 342
|||||

Search completed: May 23, 2004, 02:42:24
Job time : 81 secs

Db 5501 CGACCAGCCTTGGAAAAGTCACTGAAAAATCTTCAATTGGACTATGTTGACCTCTATCTT 5560
QY 343 GTTCACTGCCCCCATTTGCTGCCGAGAAGATGCGCAGGGTGAGCCCCAAGATTGGCCCTGAC 402
Db 5561 ATTCAATTTCCAGTGTCTGTAAAGCCAGGTGAGGAG-----TGATCCCAAAA 5608
QY 403 GSCAAATACGTCTATTCTCAAGGACCTGACCCAGAGAACCCCGAGCCCCACATGGCGCGCTATG 462
Db 5609 GATGAAAATGGAATAAATACTATTTCACACAGTGGATCTCTGTGCCACATGGGAGGCCATG 5668
QY 463 GAGAGATTATTAGGAGATCGCAAGCCAGGTCCATTGGTGTCTCCAACTGGAGCCATTGCC 522
Db 5669 GAGAGGTGTAAGATGACAGATTGGCAAGTCCATCGGGGTGTCCAATTTCAACACAGG 5728
QY 523 GACCTTGAAGAGATGTCCAAGTTTCG-----CAAGGTCAATCCCTCAGCCCAACAGATC 576
Db 5729 CTGCTGGAGATGATCTCTCAACAGCCAGGGCTCAAGTACAAGCTGTCTGCAACCCAGGTG 5788
QY 577 GAGATTCAACCCCTTCTCTGCCCAACGAGGAGCTGGTGCAGTACTGCTTCTCCAGAGACATT 536
Db 5789 GAATGTCTCTTACTTCAACACAGAGAAACTGCTGGATTCTGCAAGTCAAAAGACATT 5848
QY 637 ATGCCCGTGGCTTACTCTCTCTGGGTCTGCAGAAACCCAGTTTC-----CACACC 687
Db 5849 GTTCTGTTGCCCTATAGTCTCTGGATCCCATCGAAGAACCATGGGTGACCCGAAC 5908
QY 688 GGTGAGCGGTGAGCGAGAGAAAGACTCTGAACAGAGATCGCCGAGAGAGGGCGGCAACACC 747
Db 5909 TCCCGGTGCTCTTGGAGGAGCCAGTCTCTTGTGCTTGGCAAAAGAACACAAAGCGAAC 5968
QY 748 CTGCTCAGGTTCTTATTGCTGGGTCTGCGCGCTGCTGCTAGCTGCTTCTCCCAAGAGC 807
Db 5969 CCAGCCCTGATTGCTTGGCTTGGCTACCACTGACGCTGGGTGTGGTCTGCTGCGCAAGAGC 6028
QY 808 TCCAAACCCCAAGCGCATTTGAGTCCAACTTCAAGAGCATGAGCTC 852
Db 6029 TACAATGAGCGCGCATCAGACAGAACGTGAGGTGTTGAATTC 6073

```
QY 808 TCCAAACCCCAAGCGCATTGAGTCCAACTTCAAGAGCATTGAGCTC 852
DB 1145 TACATGAGCAGCGCATCAGACAGAACCTGCGAGGTGTTGATTC 1189

RESULT 13
US-09-702-705-784
; Sequence 784, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 784
; LENGTH: 6353
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-702-705-784

Query Match 10.9%; Score 106.6; DB 4; Length 6353;
Best Local Similarity 50.5%; Pred. No. 7e-21;
Matches 386; Conservative 0; Mismatches 349; Indels 30; Gaps 4;

QY 103 GCTGTCCACCTGCCCTGAAGCCGGTTACCGTCACCTTGGACTGTGCTGTACTACCTG 162
DB 5324 GCGGTCAAATTGGCAATAGAGCCGGTTCCACCATAATTGATCTTGACATGTTTACAAT 5383
QY 163 AACGAGGTGAGTGTGGTGAAGGTATCCGTGACCTTCTTGAAGGAGAACCCCTCGGTGAAG 222
DB 5384 ATGAGGAGCATATTTACACTTCAAAGCTTGAAGGCTTTGGAGCAATTCCTATCGACAGTTGGTC 5500
QY 283 CTCTGGTCCATTGACGACTTCCCTGAAGCGTCTTGGACTTGAAGTGTGATATGTTCTC 342
DB 5501 CGACGAGCTTGAAGGCTCACTGAAAATCTTCAATTGGACTATGTTGACCTCTATCTT 5560
QY 343 GTTCACTGGCCCAATGCTGCCAGAGAAATGGCCAGGTTGAGCCCAAGATTGGCCCTGAC 402
DB 5561 ATTCATTTTCCAGTGTCTGTAAGGCTGAGGAGTGGAGGAGTGGAGGAGTGGAGGAGTGGAG 5608
QY 403 GCGAATACGTCATTTCAAGGACCTGACCGAGAACCCGAGCCACATGCGCGGTATG 462
DB 5609 GATGAAATGGAATAACTATTTGACACATGATCTCTGTGCAATGGAGGCGCATG 5668
QY 463 GAGAGATTTATGAGGATCGAAGCCAGGTCATTTGGTGTCTCCAACTGGACCAATTGCC 522
DB 5669 GAGAGTGTAAAGATGACGAGGATTTGGCCAAAGTCCATCGGGGTGTCCTCAACTTCAACACAGG 5728
QY 523 GACCTTGAGAGATGTCGAAGTTCG-----CAGGTCTGCTCAGCCCAACCATC 576
DB 5729 CTGCTGGAGATGATCCTCAACAGCCAGGCTCAAGTACAGGCTGTCTGCAACGAGGTG 5788
QY 577 GAGATTTCAACCTTCTCTGCCCAACGAGGAGCTGGTGTGAGTACTGTTCTTCCAAAGACATT 636
DB 5789 GAAATGTCATCTTACTTCAACAGAGAAATCTGTGGATTTCTGCAAGTCAAGACATT 5848

US-09-736-457-784
; Sequence 784, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 784
; LENGTH: 6353
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-736-457-784

Query Match 10.9%; Score 106.6; DB 4; Length 6353;
Best Local Similarity 50.5%; Pred. No. 7e-21;
Matches 386; Conservative 0; Mismatches 349; Indels 30; Gaps 4;

QY 103 GCTGTCCACCTGCCCTGAAGCCGGTTACCGTCACCTTGGACTGTGCTGTACTACCTG 162
DB 5324 GCGGTCAAATTGGCAATAGAGCCGGTTCCACCATAATTGATCTTGACATGTTTACAAT 5383
QY 163 AACGAGGTGAGTGTGGTGAAGGTATCCGTGACCTTCTTGAAGGAGAACCCCTCGGTGAAG 222
DB 5384 ATGAGGAGCATATTTACACTTCAAAGCTTGAAGGCTTTGGAGCAATTCCTATCGACAGTTGGTC 5500
QY 283 CTCTGGTCCATTGACGACTTCCCTGAAGCGTCTTGGACTTGAAGTGTGATATGTTCTC 342
DB 5501 CGACGAGCTTGAAGGCTCACTGAAAATCTTCAATTGGACTATGTTGACCTCTATCTT 5560
QY 343 GTTCACTGGCCCAATGCTGCCAGAGAAATGGCCAGGTTGAGCCCAAGATTGGCCCTGAC 402
DB 5561 ATTCATTTTCCAGTGTCTGTAAGGCTGAGGAGTGGAGGAGTGGAGGAGTGGAGGAGTGGAG 5608
QY 403 GCGAATACGTCATTTCAAGGACCTGACCGAGAACCCGAGCCACATGCGCGGTATG 462
DB 5609 GATGAAATGGAATAACTATTTGACACATGATCTCTGTGCAATGGAGGCGCATG 5668
QY 463 GAGAGATTTATGAGGATCGAAGCCAGGTCATTTGGTGTCTCCAACTGGACCAATTGCC 522
DB 5669 GAGAGTGTAAAGATGACGAGGATTTGGCCAAAGTCCATCGGGGTGTCCTCAACTTCAACACAGG 5728
QY 523 GACCTTGAGAGATGTCGAAGTTCG-----CAGGTCTGCTCAGCCCAACCATC 576
DB 5729 CTGCTGGAGATGATCCTCAACAGCCAGGCTCAAGTACAGGCTGTCTGCAACGAGGTG 5788
QY 577 GAGATTTCAACCTTCTCTGCCCAACGAGGAGCTGGTGTGAGTACTGTTCTTCCAAAGACATT 636
DB 5789 GAAATGTCATCTTACTTCAACAGAGAAATCTGTGGATTTCTGCAAGTCAAGACATT 5848
```

```
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671,325
; CURRENT FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 807
; LENGTH: 3829
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-671-325-807

Query Match      10.9%; Score 106.6; DB 4; Length 3829;
Best Local Similarity 50.5%; Pred. No. 5.5e-21;
Matches 386; Conservative 0; Mismatches 349; Indels 30; Gaps 4;

QY 103 GCTGTCACCACTGCCCTGAAGACCGGTTACCGTACCTTGGACTGGCTGTACTACTCTG 162
DB 440 GCCGTCAAAATGGCAATAGAAGCCGGTTCCACCATATTGATCTGCACATGTTTACAAT 499
QY 163 AACGAGGGTGAGGTTGGTGAGGATATCCGTGACTTCTTGAAGAGAACCCCTCGTGAAG 222
DB 500 AATGAGGAGCAGGTTGGACTGGCCATCCG---AAGCAAGATTGCAGATGGCAGTGTGAAG 556
QY 223 CGTGTGACCATCTTCCGTGCAACCAAGGTGTGGAACCACTCCACCGTTATGAGGACGTC 282
DB 557 AGAAGAGACATATTCTAGACTTCAAGCTTTGGAGCAATTCCTCCACGAGTTGGTTC 616
QY 283 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 342
DB 617 CGACAGAGCTTTGGAAGGTCACTGAAAATCTTCAATTGGACTATGTTGACCTCTATCTT 676
QY 343 GTTCACTGCCCCATTTCTCAAGGACCTTGACGAGAACCCCGAGCCACATATGGCCCTGAC 402
DB 677 ATTCAATTTCCAGTGTCTGTAAGCCAGGTGAGGAG-----TGATCCCAAAA 724
QY 403 GGCAATAAGTCAATTTCTCAAGGACCTTGACGAGAACCCCGAGCCACATATGGCCGCTATG 462
DB 725 GATGAAAAATGGAATAATACTATTGACACAGTGGATCTCTGTGCCACATGGGAGGCCATG 784
QY 463 GAGAAATTTATGAGGATCGCAAGCCAGGTCATGTTGTTCTTCAACTGGACCATTTGCC 522
DB 785 GAGAAATGTAAGATCAGGATTTGGCAAGTCCATCGGGGTGTCCAACTTCAACCAAG 844
QY 523 GACTTTGAGAAGATGTCCAAAGTTTCG-----CAAGGTCACTGCTCACGCCAACACAGATC 576
DB 845 CTGCTGAGATGATCTCTCAAGCCAGGCTCAAGTACAAGCCTGTCTGCAACCAAGTG 904
QY 577 GAGATTCAACCCCTTCTGCCCCAAGGAGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 636
DB 905 GAATGTATCTCTTACTTCAACCAAGGAAACTGCTGATTTCTGCAAGTCAAAAGACATT 964
QY 637 ATCCCGCTGGCTTACTCTCTCTGCGCTCGCAGAACCAAGTTTC-----CACCAAC 687
DB 965 GTTCTGTTGCTTATAGTGTCTTGGATCCCATCGAGAGAACCATGGGTGGACCCGAC 1024
QY 688 GTGAGCGGGTCAGGAGACAGACTCTGAACGAGATCGCGAGAGGGCGGCAACAC 747
DB 1025 TCCCCGCTGCTTTGGAGAACCCAGTCTTTGTGCTTTGGCAAAAAGCAACACGCAAC 1084
QY 748 CTGTGCTAGGTTCTTATTTGCTGGGCTGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 807
DB 1085 CCAGCCCTGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1144
QY 808 TCCACCCCAAGCGATTGAGTCCAACTTCAAGACATTGAGCTC 852
DB 1145 TACAAATGACGCGCATCAGACAGAACGTGCAGGTGTTTGAATTC 1189
```

RESULT 12

US-09-589-184-807

; Sequence 807, Application US/09589184

; Patent No. 6686447

```
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darriek
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.478C8
; CURRENT APPLICATION NUMBER: US/09/589,184
; CURRENT FILING DATE: 2000-06-05
; NUMBER OF SEQ ID NOS: 827
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 807
; LENGTH: 3829
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-589-184-807
```

```
Query Match      10.9%; Score 106.6; DB 4; Length 3829;
Best Local Similarity 50.5%; Pred. No. 5.5e-21;
Matches 386; Conservative 0; Mismatches 349; Indels 30; Gaps 4;

QY 103 GCTGTGACCACTGCCCTGAAGACCGGTTACCGTACCTTGGACTGGCTGTACTACTCTG 162
DB 440 GCCGTCAAAATGGCAATAGAAGCCGGTTCCACCATATTGATCTGCACATGTTTACAAT 499
QY 163 AACGAGGGTGAGGTTGGTGAGGATATCCGTGACTTCTTGAAGAGAACCCCTCGGTGAAG 222
DB 500 AATGAGGAGCAGGTTGGACTGGCCATCCG---AAGCAAGATTGCAGATGGCAGTGTGAAG 556
QY 223 CGTGTGACCATCTTCCGTGCAACCAAGGTGTGGAACCACTCCACCGTTATGAGGACGTC 282
DB 557 AGAAGAGACATATTCTAGACTTCAAGCTTTGGAGCAATTCCTCCACGAGTTGGTTC 616
QY 283 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 342
DB 617 CGACCAAGCTTTGGAAGGTCACTGAAAATCTTCAATTGGACTATGTTGACCTCTATCTT 676
QY 343 GTTCACTGCCCCATTTCTCAAGGACCTTGACGAGAACCCCGAGCCACATATGGCCCTGAC 402
DB 677 ATTCAATTTCCAGTGTCTGTAAGCCAGGTGAGGAG-----TGATCCCAAAA 724
QY 403 GGCAATAAGTCAATTTCTCAAGGACCTTGACGAGAACCCCGAGCCACATATGGCCGCTATG 462
DB 725 GATGAAAAATGGAATAATACTATTGACACAGTGGATCTCTGTGCCACATGGGAGGCCATG 784
QY 463 GAGAAATTTATGAGGATCGCAAGCCAGGTCATGTTGTTCTTCAACTGGACCATTTGCC 522
DB 785 GAGAAATGTAAGATCAGGATTTGGCAAGTCCATCGGGGTGTCCAACTTCAACCAAG 844
QY 523 GACTTTGAGAAGATGTCCAAAGTTTCG-----CAAGGTCACTGCTCACGCCAACACAGATC 576
DB 845 CTGCTGAGATGATCTCTCAAGCCAGGCTCAAGTACAAGCCTGTCTGCAACCAAGTG 904
QY 577 GAGATTCAACCCCTTCTGCCCCAAGGAGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 636
DB 905 GAATGTATCTCTTACTTCAACCAAGGAAACTGCTGATTTCTGCAAGTCAAAAGACATT 964
QY 637 ATCCCGCTGGCTTACTCTCTCTGCGCTCGCAGAACCAAGTTTC-----CACCAAC 687
DB 965 GTTCTGTTGCTTATAGTGTCTTGGATCCCATCGAGAGAACCATGGGTGGACCCGAC 1024
QY 688 GTGAGCGGGTCAGGAGACAGACTCTGAACGAGATCGCGAGAGGGCGGCAACAC 747
DB 1025 TCCCCGCTGCTTTGGAGAACCCAGTCTTTGTGCTTTGGCAAAAAGCAACACGCAAC 1084
QY 748 CTGTGCTAGGTTCTTATTTGCTGGGCTGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 807
DB 1085 CCAGCCCTGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1144
```

Db 500 AATGAGGAGCAGTTGGAGTGGCCATCGG---AAGCAAGATTGCAGATGGCAGTGTGAAG 556
Qy 223 CQTGAGGACATCTTCGTCTGACCAAGGTGTGGAACCACTCCACCGTTATGAGGAGCTC 282
Db 557 AGAAGACATATCTACATCTCAAGCTTTGGAGCAATTCCTACGACAGAGTTGGTC 616
Qy 283 CTCTGGTCAATTGACGATCCCTGAAGCGTCTTGAGCTTGACTAGTTGATATGTTCTC 342
Db 617 CGACCAAGCCTTGAAAGGTCACTGAAAAATCTCAATTGGACTATGTTGACCTCTATCTT 676
Qy 343 GTTCACTGGCCATCTGCGGAGAGATGSCCAGGCTGAGCCCAAGATTGGCCCTGAC 402
Db 677 ATTCAATTTCCAGTCTGTAAAGCCAGGTGAGGAG-----TGATCCCAAAA 724
Qy 403 GGCATAATAGTCATCTCAAGGACCTGACCGAGAACCCCGAGCCACATGGCGCGCTATG 462
Db 725 GATGAAATGGAAAAATACTATTGACACAGTGGATCTCTGTGCCACATGGGAGGCCATG 784
Qy 577 GAGATTACCCCTTCTGCGCCAAAGGAGCTGTGTCAGTACTGCTTCTCCAGAACATT 636
Db 905 GAATGTCATCTTACTTCAACCAGAGAAAAAAGTCTGGAATTTCTGCAAGTCAAAAAGACATT 964
Qy 637 ATGCCCGTGGCTACTCTCTCTGGGCTCGAGAACCAAGTTCC-----CACCACC 687
Db 965 GTTCTGGTTCCTTATGCTTGTGGTCTGCGCGTGTGCTAGCTGCTTCTCCCAAGAGC 807
Db 1085 CCAGCCCTGATTGCGCTGCTTACAGCTGAGCGTGGGGTGTGGTCTCTGGCCAAAGGC 1144
Qy 808 TCCAAACCCCAAGCGCATTTGAGTCCCACTTCAAGAGCATTTGAGCTC 852
Db 1145 TACAATGACGCGCATCAGACAGAACCTGCAGGTGTTTGAATTC 1189

RESULT 10

US-09-614-124B-807

; Sequence 807, Application US/09614124B

; Patent No. 6630574

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Lodes, Michael A.

; APPLICANT: Fanger, Gary

; APPLICANT: Vedvick, Tom

; APPLICANT: Carter, Darrick

; APPLICANT: Retter, Marc

; APPLICANT: Mannion, Jane

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

; FILE REFERENCE: 210121.478C9

; CURRENT FILING DATE: 2001-07-11

; NUMBER OF SEQ ID NOS: 1668

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 807

; LENGTH: 3629

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-614-124B-807

Query Match 10.9%; Score 106.6; DB 4; Length 3829;
Best Local Similarity 50.5%; Pred. No. 5.5e-21;
Matches 386; Conservative 0; Mismatches 349; Indels 30; Gaps 4;

Qy 103 GCTGTCCACCACTGCCCTGAAGACCGGTTACCGTCACTTGGACTGTGCTGTGTTACTACCTG 162
Db 440 GCCGTCAAAATTGGCAATAGAGCGGGTTCCACCAATATGATTTGCAATCTTTACAAT 499
Qy 163 AACGAGGGTGAAGTTGGTGAAGGTATCCGTGACTTCTCTGAAGAGAACCCCTCGGTGAAG 222
Db 500 AATGAGGAGCAGGTTGGACTGGCCATCCG---AAGCAAGATTGCAGATGGCAGTGTGAAG 556
Qy 223 CGTGAGGACATCTTCGTCTGCAACCAAGGTGTGGAACCACTCCACCGTTATGAGGAGCTC 282
Db 557 AGAAGAGACATATTTACACTTCAAAAGCTTTGAGCAATTCCTATCGACAGAGTTGGTC 616
Qy 283 CTCTGGTCCATTGACGACTCCCTGAAGCGTCTTGGACTTGGACTGACTAGTTGATATGTTCTC 342
Db 617 CGACCAAGCCTTGAAAGGTCACTGAAAAATCTCAATTGGACTATGTTGACCTCTATCTT 676
Qy 343 GTTCACTGGCCATCTGCTGCGAGAGAAATGSCCAGGCTGAGCCCAAGATTGGCCCTGAC 402
Db 677 ATTCAATTTCCAGTGTCTGTAAAGCCAGGTGAGGAG-----TGATCCCAAAA 724
Qy 403 GGCATAATAGTCATCTCAAGGACCTGACCGAGAACCCCGAGCCACATGGCGCGCTATG 462
Db 725 GATGAAATGGAAAAATACTATTGACACAGTGGATCTCTGTGCCACATGGGAGGCCATG 784
Qy 463 GAGAAATTTATGAGATCGAAGGCCAGGTCAATTTGGTGTCTTCAACTGGACCAATTGCC 522
Db 785 GAGAAGTGTAAAGATGCAGGATTGGCCAAAGTCCATCGGGGTGTCCAACTTCAACCACAGG 844
Qy 523 GACCTTGAGAGATGTCCTCAAGTTCCG-----CAGGTCATGCTCAGCCCAACCAAGATC 576
Db 845 CTCTGGAGATGATCTTCAACAGCCAGGCTCAAGTCAAGCCTGTCTGCAACCAAGTG 904
Qy 577 GAGATTACCCCTTCTGCGCCAAAGGAGCTGTGTCAGTACTGCTTCTCCAGAACATT 636
Db 905 GAATGTCATCTTACTTCAACCAGAGAAAAAAGTCTGGAATTTCTGCAAGTCAAAAAGACATT 964
Qy 637 ATGCCCGTGGCTACTCTCTCTGGGCTCGAGAACCAAGTTCC-----CACCACC 687
Db 965 GTTCTGGTTCCTTATGCTTGTGGTCTGCGCGTGTGCTAGCTGCTTCTCCCAAGAGC 807
Db 1085 CCAGCCCTGATTGCGCTGCTTACAGCTGAGCGTGGGGTGTGGTCTCTGGCCAAAGGC 1144
Qy 808 TCCAAACCCCAAGCGCATTTGAGTCCCACTTCAAGAGCATTTGAGCTC 852
Db 1145 TACAATGACGCGCATCAGACAGAACCTGCAGGTGTTTGAATTC 1189

RESULT 11

US-09-671-325-807

; Sequence 807, Application US/09671325

; Patent No. 6667154

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Lodes, Michael A.

; APPLICANT: Fanger, Gary

; APPLICANT: Vedvick, Tom

; APPLICANT: Carter, Darrick

; APPLICANT: Retter, Marc

; APPLICANT: Mannion, Jane

; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

Qy	523	GAC	TTGAG	AAGATG	CTCAAG	TTGCG-----CAAG	GTCA	GTCTCA	CGCCAA	CCAGATC	576
Db	796	CTG	CTGAG	ATGAT	CTCTCA	CAAGCC	AGGGCTCA	AGTCA	AGGCTGT	CTCAAC	855
Qy	577	GAG	ATTCA	CCCCCT	TCCTGCC	CAACG	AGAGT	TGCTG	CAGTACT	GTTTCT	636
Db	856	GA	ATGTC	ATC	CTTACTT	CAAC	AGAGAAA	CTG	CTGATTT	CTGCA	915
Qy	637	ATG	CCCTG	GGCTACT	CTCTCT	GGGCT	TCG	CAGAAC	CCAGTTCC-----CACC	ACC	687
Db	916	GTT	CTG	TGTTGCC	TATAG	TGCT	GGGAT	CCCAT	CGAAGA	AACCAT	975
Qy	688	GGT	GAG	CGGGT	CAG	CGAGAA	CAAGACT	CTG	AAACGAGAT	CCGCCG	747
Db	976	TC	CCG	GTG	CTTTG	GAGG	ACC	CAGT	CCTTTG	TGSC	1035
Qy	748	CTT	GCTC	AG	TTTCTT	AT	TGCT	GGGTCT	CGCCG	TGCTAC	807
Db	1036	CC	AGCC	CTGA	TG	CCCTCG	CGTAC	CA	CGCTG	AGCGTGG	1095
Qy	808	TCC	AAC	CCCCA	AGCG	CATTG	AGTCC	CAACTT	CA	AGAGCAT	852
Db	1096	TAC	AATG	AG	CAGG	CGCAT	CAG	CAGAA	CGTGC	AGGTGTT	1140

RESULT 8

```

US-09-702-705-807
; Sequence 807, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 807
; LENGTH: 3829
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-702-705-807

```

RESULT 9

```

US-09-736-457-807
; Sequence 807, Application US/09736457
; Patent No. 650948
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 807
; LENGTH: 3829
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-736-457-807

```

	Query Match	10.9%	Score 106.6	DB 4	Length 3829
	Best Local Similarity	50.5%	Pred. No. 5.5e-21		
	Matches 386	Conservative	0	Mismatches 349	Indels 30
				Gaps	4
Qy	103	GCTGTACCACTGCCCTGAAGACGGGTACCGTCACTTGGACTGTGCTCGTACTACTG	162		
Db	440	GCGCTCAAAATGGCAATAGAACGGGTTCCACCATATTGATTCTGCACATGTTTACAAT	499		
Qy	163	AACGAGGGTGAGGTGTGGTGAAGGGTATCCGTGACTTCTCGAAGGAGAACCCCTCGGTGAAG	222		

QY 808 TCCAAACCCCAAGCGCATTGAGTCCAACTTCAAGAGCATTTAGCTC 852
Db 1096 TACAATGACGAGCGCATCAGACAGAACGTCGAGGTGTTTGAATTC 1140

RESULT 6

US-09-542-615A-171
; Sequence 171, Application US/09542615A

; Patent No. 6518256

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Pan, Liqun

; APPLICANT: Kalos, Michael D.

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Hosken, Nancy A.

; APPLICANT: Fanger, Gary R.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY

; FILE REFERENCE: 210121.455C8

; CURRENT APPLICATION NUMBER: US/09/542,615A

; CURRENT FILING DATE: 2000-04-14

; NUMBER OF SEQ ID NOS: 350

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 171

; LENGTH: 1491

; TYPE: DNA

; ORGANISM: Homo sapien

; US-09-542-615A-171

Query Match 10.9%; Score 106.6; DB 4; Length 1491;

Best Local Similarity 50.5%; Pred. No. 3.5e-21;

Matches 386; Conservative 0; Mismatches 349; Indels 30; Gaps 4;

QY 103 GCTGTACACATGCGCTGAAGACCGGTTACCGTCACTTGGACTGTGCTGTACTACTG 162

Db 391 GCGTCAAAATGGCAATAGAAGCCGGTTCCACCATAATTGATCTGCACATGTTTACAAT 450

QY 163 AACGAGGTGAGTTGGTGGGATTCGTCGACTTCTTGAAGGAGAACCCCTCGGTGAAG 222

Db 451 AATGAGGACGAGTTGGACTGCGCATCCG---AAGCAAGATTGCAGATGGCAGTGAAG 507

QY 223 COTGAGGACATCTTCGTCTGCACCAAGGTGTGAACCACTCCACCGTTATGAGACGTC 282

Db 508 AGAGAAGACATATCTACATCTCAAAGCTTTGGAGCAATTCCTCATCGACAGATTGTC 567

QY 283 CTCGTGTCATGACGACTCCCTGAGCGTCTTGGACTTGACTGACTGATGATGTCCTC 342

Db 568 CGACCAGCCTTGGAAGGTCACTGAAATCTTCAATTGGACTATGTTGACCTCTATCTT 627

QY 343 GTTCACTGGCCCATTTGCTGCCGAGAAGAAATGGCCAGGTGAGCCCAAGATTGGCCCTGAC 402

Db 628 ATTCATTTTCCAGTCTCTGTAAAGCCAGGTGAGGAG-----TGATCCCAAAA 675

QY 403 GGCAAATAGCTCATTTCTCAAGACCTGTACCGAGAACCCCGAGCCCAATGCGCGCTATG 462

Db 676 GATGAAATGGAATAATACTATTTGACACAGTGGATCTCTGTGCCACATGGGAGGCCATG 735

QY 463 GAGAAGATTATGAGGATGCAAGSCCAGGTCCATTTGTTGTTCTCAACTGGACCAATTGCC 522

Db 736 GGAAGTGTAAAGATGCAAGATTGGCCAAAGTCCATTCGGGGTGTCCAACTTCAACACAGG 795

QY 523 GACCTTGAAGAGATGTCGAAGTTGCG-----CAAGGTCATGCTCTCAAGCCCAAGATC 576

Db 796 CTGCTGGAGATGATCCTCAACAGCCAGGCTCAAGTACAAGCCTGTCTGCAACAGGTG 855

QY 577 GAGATTACCCCTTCTCCCAACGAGGAGTGTGTCAGTACTGCTTCTCCAGAACATT 636

Db 856 GAATGTATCTTACTTCAACACAGAGAAACTCTGTGAGATTCTTCAAGTCAAAAGACATT 915

QY 637 ATGCCCGTGGCCTACTCTCTCTGCGCTCGAGAACCAAGTTTC-----CACACC 687

Db 916 GTTCTGGTTGCTATAGTCTCTGSGATCCCATCGAGAAGAACCATGGGTGGACCCGAAC 975

QY 688 GGTGACGGGTACGCGAGAAACAGACTCTGAACGAGATCGCCGAGAGGGCGGCAACACC 747

Db 976 TCCCCGGTGTCTTTGAGGAGACCCAGTCCTTTGTGCTTGGCAAAAGACACAGGCAACC 1035

QY 748 CTTGCTCAGGTTCTTATTGCTTGGGGTCTGCGCGCTGAGCTGTAGCTGTTCTCCCCCAAGAGC 807

Db 1036 CCAGCCCTGATTGCCCTCGCTACCAAGCTGCAGCGTGGGGTGTGGTCTCTGCCAAGAGC 1095

QY 808 TCCAAACCCCAAGCGCATTTAGTCCAACTTCAAGAGCATTTAGCTC 852

Db 1096 TACAATGACGCGCATCAGACAGAACGTCGAGGTGTTTGAATTC 1140

US-09-606-421B-171

; Sequence 171, Application US/09606421B

; Patent No. 6531315

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Pan, Liqun

; APPLICANT: Kalos, Michael D.

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Hosken, Nancy

; APPLICANT: Fanger, Gary R.

; APPLICANT: Li, Samuel X.

; APPLICANT: Wang, Aijun

; APPLICANT: Skeiky, Yasir A.W.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.455C9

; CURRENT APPLICATION NUMBER: US/09/606,421B

; CURRENT FILING DATE: 2000-06-28

; NUMBER OF SEQ ID NOS: 358

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 171

; LENGTH: 1491

; TYPE: DNA

; ORGANISM: Homo sapien

; US-09-606-421B-171

Query Match 10.9%; Score 106.6; DB 4; Length 1491;

Best Local Similarity 50.5%; Pred. No. 3.5e-21;

Matches 386; Conservative 0; Mismatches 349; Indels 30; Gaps 4;

QY 103 GCTGTACACATGCGCTGAAGACCGGTTACCGTCACTTGGACTGTGCTGTACTACTG 162

Db 391 GCGTCAAAATGGCAATAGAAGCCGGTTCCACCATAATTGATCTGCACATGTTTACAAT 450

QY 163 AACGAGGTGAGTTGGTGGGATTCGTCGACTTCTTGAAGGAGAACCCCTCGGTGAAG 222

Db 451 AATGAGGACGAGTTGGACTGCGCATCCG---AAGCAAGATTGCAGATGGCAGTGAAG 507

QY 223 CGTGAGGACATCTTCGTCTGCACCAAGGTGTGAACCACTCCACCGTTATGAGACGTC 282

Db 508 AGAGAAGACATATTTACACTTCAAAGCTTTGGAGCAATTCCTCATCGACAGATTGTC 567

QY 283 CTCGTGTCATGACGACTCCCTGAGCGTCTTGGACTTGACTGACTGATGATGTCCTC 342

Db 568 CGACCAGCCTTGGAAGGTCACTGAAATCTTCAATTGGACTATGTTGACCTCTATCTT 627

QY 343 GTTCACTGGCCCATTTGCTGCCGAGAAGAAATGGCCAGGTGAGCCCAAGATTGGCCCTGAC 402

Db 628 ATTCATTTTCCAGTCTCTGTAAAGCCAGGTGAGGAG-----TGATCCCAAAA 675

QY 403 GGCAAATAGCTCATTTCTCAAGACCTGTACCGAGAACCCCGAGCCCAATGCGCGCTATG 462

Db 676 GATGAAATGGAATAATACTATTTGACACAGTGGATCTCTGTGCCACATGGGAGGCCATG 735

QY 463 GAGAAGATTATGAGGATGCAAGSCCAGGTCCATTTGTTGTTCTCAACTGGACCAATTGCC 522

Db 736 GGAAGTGTAAAGATGCAAGATTGGCCAAAGTCCATTCGGGGTGTCCAACTTCAACACAGG 795

QY 523 GACCTTGAAGAGATGTCGAAGTTGCG-----CAAGGTCATGCTCTCAAGCCCAAGATC 576

Db 796 CTGCTGGAGATGATCCTCAACAGCCAGGCTCAAGTACAAGCCTGTCTGCAACAGGTG 855

QY 577 GAGATTACCCCTTCTCCCAACGAGGAGTGTGTCAGTACTGCTTCTCCAGAACATT 636

Db 856 GAATGTATCTTACTTCAACACAGAGAAACTCTGTGAGATTCTTCAAGTCAAAAGACATT 915

QY 637 ATGCCCGTGGCCTACTCTCTCTGCGCTCGAGAACCAAGTTTC-----CACACC 687

Db 916 GTTCTGGTTGCTATAGTCTCTGSGATCCCATCGAGAAGAACCATGGGTGGACCCGAAC 975

```
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 171
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-643-597-171

Query Match      10.9%; Score 106.6; DB 4; Length 1491;
Best Local Similarity 50.5%; Pred. No. 3.5e-21;
Matches 386; Conservative 0; Mismatches 349; Indels 30; Gaps 4;

QY 103 GCTGTCAACCACTGCCCTGAAGACCGGTTACCGTCACTTGGACTGTGCTGTACTACCTG 162
    |||||
Db 391 GCCGTCAAATGGCAATAGAGCCGGTTCCACCATATTGATCTGCACATGTTTACAAT 450

QY 163 AACGAGGTGAGTGGTGGAGGATATCCGTGACTTCTTGAAGAGAACCCCTCGGTGAAG 222
    |||||
Db 451 AATGAGGAGCAGTGGGACTGGCCATCCG---AAGCAAGATTGCAGATGGCAGTGTGAAG 507

QY 223 CQTGAGGACATCTTCGTCTGCACCAAGGTGTGGAACCACTCCACCGTTATGAGGAGTC 282
    |||||
Db 508 AGAAGAGACATATCTACACTCAAGCTTTGGAGCAATTCCTACGACAGATTTGGTC 567

QY 283 CTCTCGTCCATTGACACTCCCTGAAGCGTCTTGGACTGTGACTAGTATGATGTTCTC 342
    |||||
Db 568 GCACACAGCCTTGGAAAGTCACTGAAAAATCTTCAATTGGACTATGTTGACCTTATCTT 627

QY 343 GTTCACTGGCCATTGCTGCCAGAGAAATGGCCAGGTGAGCCCAAGATTGGCCCTGAC 402
    |||||
Db 628 ATTCAATTTCCAGTGTCTGTAAGCCAGGTGAGGAAG-----TGATCCCAAAA 675

QY 403 GGCAAAATAGCTCAATTTCTCAAGGACCTGACCGAGAACCCCGAGCCCACTATGGCGCTATG 462
    |||||
Db 676 GATGAAATGGAAAAATACTATTTGACACAGTGGATCTCTGTGCCACATGGGAGGCCATG 735

QY 463 GAGAAGATTATGAGATCGAAGGCCAGGTCCATTGGTGTCTCCTCAACTGGACCATTTGCC 522
    |||||
Db 736 GAGAAGTGTAAAGATGACAGGATTGGCCAAAGTCCATCGGGGTGTCCAACCTTCAACACAGG 795

QY 523 GACCTTGAGAAGATGTCCTCAAGTTCCG-----CAAGGTCACTGCCTCACGCCAACAGATC 576
    |||||
Db 796 CTGCTGGAGATGATCTCTCAACAGCCAGGGCTCAAGTACAGCCCTGTCTGCAACAGGTG 855

QY 577 GAGATTCAACCTTCTTCCGCCAAACGAGGAGTGGTGCAGTACTGCTTCTCCAAAGACATT 636
    |||||
Db 856 GAATGTCACTCTTACTTCAACACAGAAAAAATGCTGGATTCTGCAAGTCAAAAGACATT 915

QY 637 ATGCCGTGGCTACTCTCTCTGGCTCGCAGAACCCAGGTTC-----CACCAAC 687
    |||||
Db 916 GTTCTGGTTCCTTATAGTCTCTGGATCCCATCGAAGAACCATTTGGGTGACCCGGAAC 975

QY 688 GGTGAGCGGTCAGCGAGAACAAAGACTCTGAACAGATCGCCGAGAAAGCGCGCAACACC 747
    |||||
Db 976 TCCCCGGTCTCTTGGAGGACCCAGTCTTTGTGCCCTTGGCAAAAAGACAAAGCAAGCAAC 1035

QY 748 CTTGCTCAGTTCTTATTGCTTGGGTCTGCCCGGTGCTAGTCTGCTTCTCCCAAGAGC 807
    |||||
Db 1036 CCAGCCCTGATTGCCCTCGCTACCAAGTGCAGCGTGGGGTGTGGTCTTGGCAAGAGC 1095

QY 808 TCCAAACCCCAAGCGATTGAGTCCAACTTCAAGAGCAATTGAGCTC 852
    |||||
Db 1096 TACAATGACAGCGCATCAGACAGAACTGTCAGGGTGTGTAATTC 1140
```

RESULT 5

```
US-09-480-884A-171
; Sequence 171, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C6
; CURRENT APPLICATION NUMBER: US/09/480,884A
; CURRENT FILING DATE: 2001-08-27
; NUMBER OF SEQ ID NOS: 330
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 171
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-480-884A-171
```

```
Query Match      10.9%; Score 106.6; DB 4; Length 1491;
Best Local Similarity 50.5%; Pred. No. 3.5e-21;
Matches 386; Conservative 0; Mismatches 349; Indels 30; Gaps 4;

QY 103 GCTGTCAACCACTGCCCTGAAGACCGGTTACCGTCACTTGGACTGTGCTGTACTACCTG 162
    |||||
Db 391 GCCGTCAAATGGCAATAGAGCCGGTTCCACCATATTGATCTGCACATGTTTACAAT 450

QY 163 AACGAGGTGAGTGGTGGAGGATATCCGTGACTTCTTGAAGAGAACCCCTCGGTGAAG 222
    |||||
Db 451 AATGAGGAGCAGTGGGACTGGCCATCCG---AAGCAAGATTGCAGATGGCAGTGTGAAG 507

QY 223 CGTGAGGACATCTTCGTCTGCACCAAGGTGTGGAACCACTCCACCGTTATGAGGAGTC 282
    |||||
Db 508 AGAAGAGACATATCTACACTCAAGCTTTGGAGCAATTCCTACGACAGATTTGGTC 567

QY 283 CTCTGTCTCACTGACACTCCCTGAAGCGTCTTGGACTGTGACTAGTATGATGTTCTC 342
    |||||
Db 568 CGACACAGCCTTGGAAAGTCACTGAAAAATCTTCAATTGGACTATGTTGACCTTATCTT 627

QY 343 GTTCACTGGCCATTGCTGCCAGAGAAATGGCCAGGTGAGCCCAAGATTGGCCCTGAC 402
    |||||
Db 628 ATTCAATTTCCAGTGTCTGTAAGCCAGGTGAGGAAG-----TGATCCCAAAA 675

QY 403 GGCAAAATAGCTCAATTTCTCAAGGACCTGACCGAGAACCCCGAGCCCACTATGGCGCTATG 462
    |||||
Db 676 GATGAAATGGAAAAATACTATTTGACACAGTGGATCTCTGTGCCACATGGGAGGCCATG 735

QY 463 GAGAAGATTATGAGATCGAAGGCCAGGTCCATTGGTGTCTCCTCAACTGGACCATTTGCC 522
    |||||
Db 736 GAGAAGTGTAAAGATGACAGGATTGGCCAAAGTCCATCGGGGTGTCCAACCTTCAACACAGG 795

QY 523 GACCTTGAGAAGATGTCCTCAAGTTCCG-----CAAGGTCACTGCCTCACGCCAACAGATC 576
    |||||
Db 796 CTGCTGGAGATGATCTCTCAACAGCCAGGGCTCAAGTACAGCCCTGTCTGCAACAGGTG 855

QY 577 GAGATTCAACCTTCTTCCGCCAAACGAGGAGTGGTGCAGTACTGCTTCTCCAAAGACATT 636
    |||||
Db 856 GAATGTCACTCTTACTTCAACACAGAAAAAATGCTGGATTCTGCAAGTCAAAAGACATT 915

QY 637 ATGCCGTGGCTACTCTCTCTGGCTCGCAGAACCCAGGTTC-----CACCAAC 687
    |||||
Db 916 GTTCTGGTTCCTTATAGTCTCTGGATCCCATCGAAGAACCATTTGGGTGACCCGGAAC 975

QY 688 GGTGAGCGGTCAGCGAGAACAAAGACTCTGAACAGATCGCCGAGAAAGCGCGCAACACC 747
    |||||
Db 976 TCCCCGGTCTCTTGGAGGACCCAGTCTTTGTGCCCTTGGCAAAAAGACAAAGCAAGCAAC 1035

QY 748 CTTGCTCAGTTCTTATTGCTTGGGTCTGCCCGGTGCTAGTCTGCTTCTCCCAAGAGC 807
    |||||
Db 1036 CCAGCCCTGATTGCCCTCGCTACCAAGTGCAGCGTGGGGTGTGGTCTTGGCAAGAGC 1095
```


Best Local Similarity 50.8%; Pred. No. 7e-25;
Matches 435; Conservative 0; Mismatches 388; Indels 33; Gaps 5;
QY 5 CTAACGGAAAGACTTTCATATGACACACGGGTCAAGATTCCTGGCTCGGCTTTGGTA 64
Db 12 CCATGGCAAGCGCTCTCTGCTCAACAACGGCGCAAGATCCCATCTGGGGTTGGTA 71
QY 65 CTTTCGCTAGTGAAGGTTCCRAAGGGCGAGACCTATCTGCTGTACCACTGCCCTGAAGA 124
Db 72 CTT-----GGAAGTCCCTCCAGGGCAGGTGACTGAGCGCTGAGGTGCCAATTGAG 125
QY 125 CCGGTACCGTCACTTGAAGTGTGCTGCTGCTACTACTGAAACGAGGGTGAAGTTGGTAGG 184
Db 126 TCGGTACCGGCACATCGACTGTGCCATGTGTACCAAGTGAAGTGAAGTGGGGTGG 185
QY 185 GTATCCGTGACTTCTGTAAGGAGACCCCTCGGTGAAGCGTGAAGACATCTCTCTGCA 244
Db 186 CATTCAGAGAACTCAGGGAGAGGT---GGTGAAGCGTGAAGAGCTCTTCATCGTCA 242
QY 245 CCAAGTGTGGAACCACTCCACCGTTATGAGGAGCTCTCTGTGCTCAATTGACGACTCCC 304
Db 243 GCAAGCTGTGGTGACGTACCATGAGAGGGCGCTGGTGAAGGAGCTGCCAGAGACAC 302
QY 305 TGAAGCGTCTTGGACTTGAAGTGTGATATGTTCTCTGTTCACTGGGCCATTTGCTGCCG 364
Db 303 TCAGCGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 356
QY 365 AGAAGATGGCCAGGCTGAGCCCAAGATTGGCCCTGAGCGCAAAATACGTCTATTCTCAAGG 424
Db 357 -----GCTTTAAGCTGGAGAGATTTTCCCATTTGATGAGTCGGGCAATGTTTC 410
QY 425 ACTGACGAGAAACCCGAGCCCATGATGGGGCTATGGAGAGATTTATGAGATGCA 484
Db 411 CCAAGTGACACCAACATTCGACACGCTGGGGCCCATGGAAGAGCTGGTGAAGAGGC 470
QY 485 AGGCCAGGTTCATTTGGTGTCTCCAACTGGACCACTTGGCCCTGAGCGCAAAATACGTCTATTCTCAAGG 542
Db 471 TGGTGAAGCTATTGGCATCTCCAACTTCAACCATCTCCAGTGGAGATGATCTTAACA 530
QY 543 ----GTTCCGCAAGTGTATGCTTCTCCAGAACATTTATCCCGTGGCTTCTCTC 598
Db 531 AACCTGGGTTGAAGTATGAAGCTGAGCTGAGTAAACCAAGATTGAGTGCACCATATCTCACTC 590
QY 599 AGGAGAGCTGTGAGTACTGCTTCTCCAGAACATTTATCCCGTGGCTTCTCTC 658
Db 591 AGGAGAGTAAATCCAGTACTCCAGTCCAAAGGATCTGTTGACCCCTTACAGCCCCC 650
QY 659 TGGGCTCGCAGAACAGGTTCCCAACCCGCTGAG-----CGGGTCAAGCAGAACAGA 712
Db 651 TCGGCTCTCTGACAGGCTTGGGCGCAAGCCGAGGACCTTCTCTCTGAGGATCCCA 710
QY 713 CTCTGAACGAGATCCGAGAGAGGGCGGACACCCCTTGTCTGAGTCTTATTTGCTGGG 772
Db 711 GGATCAAGGCGATCGCAGCAAGCAATAAACTACAGCCAGGTCTGATCCGGTTC 770
QY 773 GTCCTGGCGGTGGTACGTCTCTCCCAAGAGCTCCAAACCCAGGCAATTCAGTCCA 832
Db 771 CATTCAGAGAACTTGTGTGTATCCCAAGTCTGTGACACCAAGCAAGCAATTCGTGAGA 830
QY 833 ACTTCAAGAGCATGA 848
Db 831 ACTTTAAGTCTTTGA 846

RESULT 2

US-08-801-344-3
; Sequence 3, Application US/08801344
; Patent No. 6087140
; GENERAL INFORMATION:
; APPLICANT: Camerton, Douglas C.
; APPLICANT: Shaw, Anita J.
; APPLICANT: Altaras, Nedim E.
; TITLE OF INVENTION: MICROBIAL PRODUCTION OF 1,2-PROPANEDIOL

; TITLE OF INVENTION: FROM SUGAR
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Demitt Ross & Stevens S.C.
; STREET: 8000 Excelsior Drive, Suite 401
; CITY: Madison
; STATE: WI
; COUNTRY: U.S.A.
; ZIP: 53717-1914
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,344
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sara, Charles S.
; REFERENCE/DOCKET NUMBER: 09820.037
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-831-2100
; TELEFAX: 608-831-2100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1337 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Rat lens aldose reductase gene
; US-08-801-344-3

Query Match 11.6%; Score 113; DB 3; Length 1337;
Best Local Similarity 50.7%; Pred. No. 4.5e-23;
Matches 424; Conservative 0; Mismatches 380; Indels 33; Gaps 5;
QY 24 ATTGAGCAACGGCGTCAAGATTCCTGGCGCTTGGTACCTTCGTTAGTGAAGGTTTC 83
Db 53 ACTCAACACGGACCAAGATGCCACCTGGTCTGGCACCT-----GGAAGTCTCC 106
QY 84 CAAGGGCGAGACCTATCTGCTGTACCACTGCCCTGAAGACCGTTACCGTCACTTGA 143
Db 107 TCCTGGCCAGGTGACCGAGGCTGTGAAGGTTGCTATCGACATGGGGTATCGCCACATTGA 166
QY 144 CTGTGCTGTACTACTGAAAGGGTGAAGTGGTGAAGGTTATCCGTGACTTCTCTGAA 203
Db 167 CTGCGCCCAAGTGTACCAAGATGAGAGAGGTGGGGTGGCCCTCCAGGAGAGCTCAA 226
QY 204 GGAGAACCCCTCGGTGAAGCGTGAAGACATCTTCTGCTGACCAAGGTTGGAACCACT 263
Db 227 GGAGCAGGT---GGTGAAGCGCAGGATCTCTTCTTATGTCAGCAAGCTGTGGTCACTG 283
QY 264 CCACCGTTATGAGGACGTCTCTGCTGTCCATTGACGACTCCCTGAAGCGCTTTGAGATTGA 323
Db 284 CCAGACACGAGATGTTGAAGGGGCTGCGCAAGAGACGCTGAGCGACCTGCAGCTGA 343
QY 324 CTACGTTGATATGTTCTGCTTCACTGGGCCATTGCTGCCGAGAAAGATGGCCAGGTTGA 383
Db 344 CTACCTGGACCTCTACCTTATTCACCTGGCCAACTG-----GCTTCAAGGCTCG 391
QY 384 GCCCAAGATTGGCCCTGACGGCAATACGTCATCTTCAAGGACCTGACCGAGAACCCCA 443
Db 392 GCCTGACTATTTCCTCCCTGGATGCAATCGGAAACGTTATCTCTAGTACACCGATTTTGT 451
QY 444 GCCCAGATGGCGCTATGAGAGATTTATGAGGATCGCAAGGCGAGGTCCATTGCTGT 503
Db 452 GGCACATTGGACGGCTATGAGCAACTAGTGTAGAGTTTGGTAAAGCAATCGAGT 511

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2004, 01:05:39 ; Search time 76 Seconds
(without alignments)
7141.342 Million cell updates/sec

Title: US-10-004-115B-2

Perfect score: 978
Sequence: 1 atgtctaaggaagacttt.....ccaagaacctgtctgcgtga 978

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCUTS COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119.2	12.2	1335	4	US-09-023-655-1010
2	113	11.6	1337	3	Sequence 1010, Ap
3	113	11.6	1337	4	Sequence 3, Appli
4	106.6	10.9	1491	4	Sequence 3, Appli
5	106.6	10.9	1491	4	Sequence 171, App
6	106.6	10.9	1491	4	Sequence 171, App
7	106.6	10.9	1491	4	Sequence 171, App
8	106.6	10.9	1491	4	Sequence 171, App
9	106.6	10.9	3829	4	Sequence 807, App
10	106.6	10.9	3829	4	Sequence 807, App
11	106.6	10.9	3829	4	Sequence 807, App
12	106.6	10.9	3829	4	Sequence 807, App
13	106.6	10.9	3829	4	Sequence 807, App
14	106.6	10.9	3829	4	Sequence 807, App
15	106.6	10.9	3829	4	Sequence 807, App
16	106.6	10.9	3829	4	Sequence 807, App
17	106.6	10.9	3829	4	Sequence 807, App
18	106.6	10.9	3829	4	Sequence 807, App
19	103.8	10.6	1265	4	Sequence 784, App
20	103.8	10.6	1265	4	Sequence 784, App
21	103.8	10.6	1265	4	Sequence 784, App
22	103.8	10.6	1265	4	Sequence 784, App
23	103	10.3	1199	4	Sequence 784, App
24	103	10.3	1199	4	Sequence 784, App
25	101	10.1	1316	4	Sequence 784, App
26	99.2	10.1	1316	4	Sequence 784, App
27	99.2	10.1	1316	4	Sequence 784, App

28 99.2 10.1 1316 4 US-09-614-124B-323 Sequence 323, App
29 99.2 10.1 1316 4 US-09-671-325-323 Sequence 323, App
30 99.2 10.1 1316 4 US-09-589-184-323 Sequence 323, App
31 98.8 10.1 1207 3 US-09-166-412-1 Sequence 1, Appli
32 98.8 10.1 1207 4 US-08-731-320B-1 Sequence 1, Appli
33 98.8 10.1 1207 4 US-09-166-438-1 Sequence 1, Appli
34 95 9.7 1206 3 US-08-532-896-1 Sequence 1, Appli
35 83.6 8.5 1132 4 US-09-626-002-18 Sequence 18, Appl
36 79.2 8.1 1233 3 US-08-853-839-1 Sequence 3, Appli
37 78.2 8.0 1259 3 US-09-166-412-3 Sequence 3, Appli
38 78.2 8.0 1259 4 US-08-731-320B-3 Sequence 3, Appli
39 78.2 8.0 1259 4 US-09-166-438-1 Sequence 3, Appli
40 77 7.9 954 2 US-08-336-198C-2 Sequence 2, Appli
41 77 7.9 954 4 US-09-184-965-2 Sequence 2, Appli
42 75.2 7.7 1073 3 US-09-347-803-5 Sequence 5, Appli
43 72.8 7.4 530 4 US-09-702-705-87 Sequence 87, Appl
44 72.8 7.4 530 4 US-09-736-457-87 Sequence 87, Appl
45 72.8 7.4 530 4 US-09-614-124B-87 Sequence 87, Appl

ALIGNMENTS

RESULT 1

US-09-023-655-1010
; Sequence 1010, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HEREMITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1010:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1335 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g179035
US-09-023-655-1010

Query Match

12.2%; Score 119.2; DB 4; Length 1335;

Db 532 GAGGAAGCTTCGGGCAAGCTTCCTCGGTAAACAGATCGAAGCTCACCCCTTGTGCAA 591
Qy 598 AACGAGAGCTGTCAGTACTGCTTCTCCAGAACATTAATGCGGTGGCTACTCTCCT 657
Db 592 CAACAGAGTGGTGTCTCACCATAAGAGCAAGAACATGTGCATCTGCTTACAGTCCC 651
Qy 658 CTGGCTCGCAGAACACAGGTTCCACCACCGGTGAGCGGTGACGAGAACAAAGACTCTG 717
Db 652 TTGGGCAACAATGTGCTGGTAAACCACTC-----TGACTGAGATCCCGGTATT 702
Qy 718 AACGAGATCGCGGAGAGGGGGCAACACCTTGTCTAGGTTCTTATGCTGGGTCTG 777
Db 703 GTGATGCTGTAAAGCTCTGAACCATCTCTGCTGTGCTCATGCTTGGGGTATT 762
Qy 778 CGCGTGTCTAGCTGCTTCTCCCAAGAGCTCCAAACCCCAAGCGCAATTGATCCAACTTC 837
Db 763 CAACGGGTACAGGCTCTTGGTCAAGTCAAGTACAGCTTACACCTCAGGATTAGAGTAACTTT 822
Qy 838 AAGAGCATGAGCTCTCGGATGCGGACTTTGAAGCCATCAATGCGGTGGCAAGG 892
Db 823 GAACAGATCACTGTCTGTATGAGGAATTCACAGGGTTACCAACCTCATCAAGG 877

RESULT 13
US-10-425-114-17144
; Sequence 17144, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425.114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 17144
; LENGTH: 1289
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3068-021-G2_FLI
US-10-425-114-17144

Query Match 13.4%; Score 130.8; DB 13; Length 1289;
Best Local Similarity 52.4%; Pred. No. 3.4e-31;
Matches 416; Conservative 0; Mismatches 357; Indels 21; Gaps 5;
Qy 107 TCACACTGCTGAGACCGGTACCGTCACTTGGACTGCTGCTGCTACTACTGAAAG 166
Db 185 TCCACAGCGCTCCGCTCGGTACCGCCACCTGGACTGCGCGCTGACTTACAGAACG 244
Qy 167 AGGTGAGTGGTGAAGGTTCCGTGCTTCTCCAGAGAACCCCTCGGTGAAGCGTG 226
Db 245 AAGCTGAGTTGGTGAACGCTCGCAGAGGCAATCCAGACCGGACTC---GTCAAGCGG 301
Qy 227 AGGACATCTTGTCTGCAACCAAGGTGGAACCACTCCACCGTTATGAGACGCTCTCT 286
Db 302 AGGACCTTTTCATCACACCAAGCTGTGGAACCTCAGACC-----ATGGCCATGTGTTG 355
Qy 287 GGTCCATTGACGACTCCCTGAAGCGTCTTGACTTGAAGTGTGATGTTCTCTGCTTC 346
Db 356 AAGCTGCAAGGACAGCCTGAAGAGCTGACGTAGACTATCTGACCTCTACCTCACTC 415
Qy 347 ACTGCCCATTTGCTGCGAGAGAAATGCGCAGGTGAGCCCAAGATTGGCCCTGACGGCA 406
Db 416 ATTTCCAGTAGCCACTCGGCACACCGGAGTCCGCACGACTCTTAGCGCTCTGGGTGACG 475
Qy 407 AATAGTCACTTCAAGGACCTGACCGGAGAACCCCGGACCCACATGCGCGCTATGAGA 466

Db 476 ATGGCGTGTGACATCGACACCACTATCTCCTCGAAGACGATGCGACCGATGGAAG 535
Qy 467 AGATTATAGAGATCGAAGGCCAGGTCCATTGGTGTCTCCACTGGACCACTTGCAGCC 526
Db 536 AGCTTGTGTTCCATGGGGTGTGCGCAGCATTTGGAATCAGCAACTACGACATCTTCTCA 595
Qy 527 TTGAGAAGATGTCCTAAGTTTCGCCAAGGTTCATGCTTCAGCCCAACAGATCGAGATTCA 586
Db 596 CCAGAGACTGCTGGCTAGCCCAAGATAAAGCGCGGTGAACAGATCGAGACGAC 655
Qy 587 CTTCTGCTCCCAACAGAGAGCTGGTGCAGTACTGC---TTCTCAAGAACATATTGCGCG 643
Db 656 CTTACTTCAGCGCGACTCTCTTGTCAAGTCTGCGCAAGACACGGGATCTGCGTGACCG 715
Qy 644 TGGCTTACTCTCTCTGGGCTCGCAGAACCAAGTTCCACACCGGTGAGCGGGTCAGCG 703
Db 716 CGCACACCGCTGGTGGTCCACCGCAACCGAGTGGTTCGGGAGCGGTCTCGTGC 775
Qy 704 AGAACAAAGACTCTGAACGAGA---TCGCGGAGAGGGCGGCAACCCCTTGTCTAGGTT 760
Db 776 TGAAGACCTCTCATCAAGAGCTTTGGCTGAGAGTAGCGGCAAGACGCGCGCGAGCTCG 835
Qy 761 TTATTGCTGGGTCTGCGCGTGGCTAGTCTGTTCTCCCAAGAGCTCCACCCCAAGC 820
Db 836 TCTCGCTGGGTCTGACGCGGACACGCTGGTATCCCAAGACCTCCCAAGCGGAGA 895
Qy 821 GCATTGAGTCCAACT-----TCAAGAGCATTTGAGCTCTCGATCCGACTTTGAAGCCA 874
Db 896 GGCTGCAGGAGAACTTTGACGCTTTTCGGTTTTCACATCTCCGCGGAGGACATGGAGAGA 955
Qy 875 TCAATGCGGTGTC 888
Db 956 TGAAGCGCGTGCAC 969

RESULT 14
US-10-425-114-25069
; Sequence 25069, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425.114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 25069
; LENGTH: 1289
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3689-245-C12_FLI
US-10-425-114-25069

Query Match 13.2%; Score 129.4; DB 13; Length 1289;
Best Local Similarity 51.8%; Pred. No. 9.5e-31;
Matches 398; Conservative 0; Mismatches 356; Indels 15; Gaps 4;
Qy 107 TCACACTGCTGAGACCGGTACCGTCACTTGGACTGCTGCTGCTACTACTGAAAG 166
Db 180 TCCACACAGCTTCGCGTCCGCTACCGCCACTCGACTGCGCGCTGACTTACCAAGAG 239
Qy 167 AGGTGAGTGGTGAAGGTTATCCGTGATCTTCTGAAGGAGAACCCCTCGGTGAAGCGTG 226
Db 240 AAGCTGAGTGGTGAAGCGCTCGCAGAGGCAATCCAGACCGGACTC---GTCAAGCGG 296

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1121
; TYPE: DNA
; ORGANISM: Trichosporonoides megachiliensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(987)
; OTHER INFORMATION:
US-10-040-416-5

Query Match      14.7%; Score 144.2; DB 14; Length 1121;
Best Local Similarity 50.9%; Pred. No. 1.6e-35;
Matches 456; Conservative 0; Mismatches 418; Indels 21; Gaps 4;

QY 1 ATGCTTAACGGAAGAGACTTTCACATTTAGAGCAACGGCTCAAGATTCCTGCGGTGCGGTTT 60
DB 1 ATGCTTACAACAAGACATCCCTCTCAACGACGGTAACCTCCATCCCTGCGCTTGGCTAC 60
QY 61 GGTACCTTCGCTAGTGAAGGTTCCAAAGGGGAGACCTATATCTGCTGTCAACCACTGCCCTG 120
DB 61 GGTACCTTCGCTAGTGAAGGTTCCAAAGGGGAGACCTATATCTGCTGTCAACCACTGCCCTG 114
QY 121 AAGACCGGTTACCTTCACTTGGACTGTGCTGGTACTTACCTGTAACGAGGTTGAGTTGGT 180
DB 115 AAGCTGGCTACCTTCACTTGGACTGTGCTGGTAAAGTTTACCAGAACCAACCGAGATTGGC 174
QY 181 GAGGTATCCGTGACTTCTGAGGAGAAACCCCTCGGTGAGCGTGAGGACATCTTCGTC 240
DB 175 CAAGCTCTCAAGGAACCTGTTTGAAGGGTGT---GTCAGCGTGAGGACCTTTTCATC 231
QY 241 TGCACCAAGGTGTGGAACCCCTCAACCGTTATAGAGGAGCTCTCTGGTCCATTCAGCAC 300
DB 232 ACTTCCAGCTTGGAAACCGCCACGCTCTCTGAGCAGCTTGAGCTGCGCTCGACGAG 291
QY 301 TCCCTGAAGCGCTTGGACCTTGACTACGTTGATATGTTCTCGTTCACTGGCCCAATTGCT 360
DB 292 ACTCTTAAGGAGTTGGTCTATCTCTATTGACCTGTACTTGAATTCATCTGGCTTGGT 351
QY 361 GCCGAGAAGAATGGCCAGGGTGA---GCCAAGATTGGCCCTGACGGCAAAATACGTCAAT 417
DB 352 TTCAAGTTCACTACTCCCGATGAAGTCTCTGCTGACCCCTTACCACCAAGGACCTTGCC 411
QY 418 CTCAGGACCTGACAGAACCCCGAGCCCATGCGCGGTATGAGAGAGATTTATGAG 477
DB 412 TACATGAGCATTCGGTCAAAATGTCGACACCTCGAAGGCGGTCTGCGCCCTGAAAAG 471
QY 478 GATCGCAAGCCAGGTCCTCATTTGGTGTCTCCAACTGGACCAATTCGCCAGCTTTGAGAGATG 537
DB 472 ACGGGTAAGAACCAAGTCGGTTGGTGTTCGAACTTCAGACATTCGCTCTGGTTCGACTT 531
QY 538 TCCAGTTCCGCCAAGGTCAATGCTCAGCCACAGATCGAGATTCACCCCTTCTGCCC 597
DB 532 ACTTCCAAAGCTTGGAAACAAACCGCCACGCTCTCTGAGCAGCTTGAGCCCTGCGCTGACGAG 291
QY 301 TCCCTGAAGCGCTTGGACCTTGACTACGTTGATATGTTCTCGTTCACTGGCCCAATTGCT 360
DB 292 ACTCTTAAGGAGTTGGTCTATCTCTATTGACCTGTACTTGAATTCATCTGGCTTGGT 351
QY 361 GCCGAGAAGAATGGCCAGGGTGA---GCCAAGATTGGCCCTGACGGCAAAATACGTCAAT 417
DB 352 TTCAAGTTCACTACTCCCGATGAAGTCTCTGCTGACCCCTACCACAGGATCTTGCC 411
QY 418 CTCAGGACCTGACCGAAGACCCCGAGCCCATGCGCGCTATGAGAGAGATTTATGAG 477
DB 412 TACGTTACGATTCGGTCAAAATGTCGACACCTCGAAGGCGGTCTGCGCCCTGAAAAG 471
QY 478 GATCGCAAGCCAGGTCCTCATTTGGTGTCTCCAACTGGACCAATTCGCCAGCTTTGAGAGATG 537
DB 472 ACGGGTAAGAACCAAGTCGGTTGGTGTTCGAACTTCAGACATTCGCTCTGGTTCGACTT 531
QY 538 TCCAGTTCCGCCAAGGTCAATGCTCAGCCCAACCGATTCGAGATTCACCCCTTCTGCCC 597
DB 532 GAGGAAGCTTCGGCGGAACGCTCTCGGTAAACCGATTCGAAGCTCACCCCTTGTTCGA 591
QY 598 AACGAGGCTGTGCGAGTACTGCTCTCCAAAGCAATATGCGCGTGGCTACTCTCCCT 657
DB 592 CAAGACGAGTTGGTGTCTCACCATAAGACGAAGAACATTTGTCATCTGCTTACAGTCCC 651
QY 658 CTGGCTCGCAGAACACGAGTTCCCAACCGCTGAGCGGGTTCAGCGAGAACAAAGACTCTG 717
DB 652 TTGGCAACATATGCTGTTGAACCACTC-----TGACTGAGAACCCCGGTATT 702
QY 718 AACGAGATCGCGAAGAGGGGGGCAACACCTTCTCTCAGGTTCTTATGCTGGGTCG 777
DB 703 GTGATGCTGTAAGCGGTTTGAACCATACTCTGCTGTGCTGCTCAATTTGCTGGGTAAT 762
QY 778 CGCGTGTGCTAGCTGCTTCTCCCAAGAGCTCCAAACCCCAAGCGCATGAGTCCAACTC 837
DB 763 CAACGGGGTACAGGCTTGGTCAAGTCAGTTACACCCCTCACGGATCAAGAGTAACCTT 822
QY 838 AAGAGCATTTGAGCTCTCGATTCGCGACTTTGAAGCCATCAATTCGCTTGCACAGG 892
```

DB 823 GAACAGATCACTCTGCTGATGAGGAATTCACAAACGGTTACCAACCTCATCAAGG 877

RESULT 12

```
US-10-040-416-3
; Sequence 3, Application US/10040416
; Publication No. US20020160480A1
; GENERAL INFORMATION:
; APPLICANT: OKURA, TETUYA
; APPLICANT: KASUMI, TAKAFUMI
; APPLICANT: ASABA, EIJI
; TITLE OF INVENTION: ERYTHROSE REDUCTASE, ITS cDNA AND CELL WHICH THE cDNA EXPRESS
; FILE REFERENCE: 217969USOXCIP
; CURRENT APPLICATION NUMBER: US/10/040,416
; CURRENT FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: US 09/800,487
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: JP 2001-001294
; PRIOR FILING DATE: 2001-01-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1077
; TYPE: DNA
; ORGANISM: Trichosporonoides megachiliensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(987)
; OTHER INFORMATION:
US-10-040-416-3

Query Match      14.6%; Score 142.6; DB 14; Length 1077;
Best Local Similarity 50.8%; Pred. No. 5.1e-35;
Matches 455; Conservative 0; Mismatches 419; Indels 21; Gaps 4;
```

```
QY 1 ATGCTTAACGGAAGACTTTCACATTTAGAGCAACGGCTCAAGATTCCTGCGGTGCGGTTT 60
DB 1 ATGCTTACAACAAGAACATCCCTCTCAACGACGGTAACCTCCATTCCTGCTTGGGTAC 60
QY 61 GGTACCTTCGCTAGTGAAGGTTCCAAAGGGGAGACCTATCTGCTGTCAACCACTGCCCTG 120
DB 61 GGTACCTTCGCTAGTGAAGGTTCCAAAGGGGAGACCTATCTGCTGTCAACCACTGCCCTG 114
QY 121 AAGACCGGTTACCTTCACTTGGACTGTGCTGGTACTTACCTGTAACGAGGTTGAGTTGGT 180
DB 115 AAGGTGGCTACCTGCTCAATTTGGACTTGGCCAAAGTATACCGAACCAACCGAGATTGGC 174
QY 181 GAGGTATCCGTGACTTCTCTGAAGGAGAAACCCCTCGGTGAAGCGTGAGGACATCTTCGTC 240
DB 175 CAAGCTCTCAAGGAACCTGTTTGAAGGGTGT---GTCAGCGTGAGGACCTTTTCATC 231
QY 241 TGCACCAAGGTGTGGAACCCCTCAACCGTTATAGAGGAGCTCTCTGGTCCATTCAGCAC 300
DB 232 ACTTCCAAAGCTTGGAAACAAACCGCCACGCTCTCTGAGCAGCTTGAGCCCTGCGCTGACGAG 291
QY 301 TCCCTGAAGCGCTTGGACCTTGACTACGTTGATATGTTCTCGTTCACTGGCCCAATTGCT 360
DB 292 ACTCTTAAGGAGTTGGTCTATCTCTATTGACCTGTACTTGAATTCATCTGGCTTGGT 351
QY 361 GCCGAGAAGAATGGCCAGGGTGA---GCCAAGATTGGCCCTGACGGCAAAATACGTCAAT 417
DB 352 TTCAAGTTCACTACTCCCGATGAAGTCTCTGCTGACCCCTTACCACCAAGGACCTTGCC 411
QY 418 CTCAGGACCTGACAGAACCCCGAGCCCATGCGCGGTATGAGAGAGATTTATGAG 477
DB 412 TACATGAGCATTCGGTCAAAATGTCGACACCTCGAAGGCGGTCTGCGCCCTGAAAAG 471
QY 478 GATCGCAAGCCAGGTCCTCATTTGGTGTCTCCAACTGGACCAATTCGCCAGCTTTGAGAGATG 537
DB 472 ACGGGTAAGAACCAAGTCGGTTGGTGTTCGAACTTCAGACATTCGCTCTGGTTCGACTT 531
QY 538 TCCAGTTCCGCCAAGGTCAATGCTCAGCCACAGATCGAGATTCACCCCTTCTGCCC 597
```

; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: JP 2000-372704
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: JP 2001-006144
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: JP 2001-026594
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: JP 2001-175175
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 18
; LENGTH: 331
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (14)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-10-004-115A-18

Query Match 27.1%; Score 265; DB 15; Length 331;
Best Local Similarity 100.0%; Pred. No. 1.4e-74;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCTAACGAAAGACTTTTACATTGAGCAACGGGCTCAAGATTCTTGGCGTCGGCTTT 60
DB 67 ATGCTAACGAAAGACTTTTACATTGAGCAACGGGCTCAAGATTCTTGGCGTCGGCTTT 126
QY 61 GGTACTTTCGGTGTAGTGAAGTTTCAAGGGCGAGACCTTACTGTGTCAACACTTGCCTG 120
DB 127 GGTACTTTCGGTGTAGTGAAGTTTCAAGGGCGAGACCTTACTGTGTCAACACTTGCCTG 186
QY 121 AAGACCGGTTACCGTCACTTGGACTGTGCTGTGTTACTTCAAGAGGGGTGAGTTGGT 180
DB 187 AAGACCGGTTACCGTCACTTGGACTGTGCTGTGTTACTTCAAGAGGGGTGAGTTGGT 246
QY 181 GAGGTAATCCGTGACTTCTCAAGGAGAACCCCTCGGTGAGCGTGAGGACATCTTCGTC 240
DB 247 GAGGTAATCCGTGACTTCTCAAGGAGAACCCCTCGGTGAGCGTGAGGACATCTTCGTC 306
QY 241 TGCACCAAGGTGTGGAACCACTCC 265
DB 307 TGCACCAAGGTGTGGAACCACTCC 331

RESULT 10
US-09-734-237B-31
; Sequence 31, Application US/09734237B
; Publication No. US20030064432A1
; GENERAL INFORMATION:
; APPLICANT: Rozzell, J. David
; APPLICANT: Bui, Peter
; APPLICANT: Hua, Ling
; TITLE OF INVENTION: SYNTHETIC GENES FOR ENHANCED EXPRESSION
; FILE REFERENCE: B583:40608
; CURRENT APPLICATION NUMBER: US/09/734,237B
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 09/494,921
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 31
; LENGTH: 972
; TYPE: DNA
; ORGANISM: Sporidiobolus salmonicolor
; US-09-734-237B-31

Query Match 20.4%; Score 199.6; DB 13; Length 972;
Best Local Similarity 55.5%; Pred. No. 2.3e-53;
Matches 463; Conservative 0; Mismatches 344; Indels 27; Gaps 3;
QY 87 GGGCGAGCTACTACTGTGTGTCACCACTGCCCTGAGACCGGTTACCGTCACTTGGACTG 146

DB 75 GGGCGAGGTGGCCAGGGCGTCAAGGTGCCATCGAGACTCGATACCGTCACTCGACT 134
QY 147 TGCCTGTACTACTTGAACGAGGTGAGGTGGTGGAGGGTATCCGTGACTTCTCTGAAAGA 206
DB 135 TGCCAAGGTCTACTCGAACCAACCTGAGGTGGTGGCCGCCATCA-----A 179
QY 207 GAACCCCTCGGTGAAGCGTGAAGGACATCTTCTGTCACCAAGGTGTGGAACCACTCCA 266
DB 180 GGAGGCTGGCTCAAGCGGAGGACCTCTTATCACTCTGAGCTCTGGAACAATCGCA 239
QY 267 CCGTTATGAGGACGTCTCTGTGTCATTGAACGATCTCTGAAAGCGTCTTGGACTTGA 326
DB 240 CCGCCCGAGGACGTGAGCTCGCTTTCGACACACCTCAAGGAGCTCGGCTCGAGTA 299
QY 327 CGTTGATATGTTCTCGTTTCACTGGCCCATCTGCTCCGAGAGAAATGGCCAGGGTGAGCC 386
DB 300 CTCGACCTTTACCTTCACTTGGCCCGTGGCTTCCGCCGAGGGGAGCATCACCCA 359
QY 387 CAAGATTGGCCCTGACGCAAAATACGTCTTCTCAAG---GACCTGACCGAAGACCCCA 443
DB 360 GAACCTCTTCCGAAAGGCCAACGACAGGAGGTCAAGCTCGACCTGGAGGTCAAGCTCGT 419
QY 444 GCGGCATGCGCGCTGTGAGAGATTTATGAGGATCGCAGGCGAGTCCATTTGGTGT 503
DB 420 CGACAGGTGAAGGGCGATGGTCAAGCTTCTGCACACTGGCAAGGTCAAGGGCGATCGGCT 479
QY 504 CTCCAACTGGACCATTTGCCGACCTTGAGAAAGATGTCGAAGTTTCGCAAGGTTCATGCC 563
DB 480 TTCCAACTTCGACGCAAGATGGTGCAGCCATCATCGAGGCTACGGCGGTGACCCCTC 539
QY 564 GCGCAACGAGATCGAGATTCAACCTTCTGCTGCCAACGAGGAGCTGGTGCAGTACTGCTT 623
DB 540 CGTCAACCAAGATCGAGCGTCAACCTTCTCTCAGGCCGAGCTCATCGCCCAACCA 599
QY 624 CTCCAAGAAATATTGCGCGTGGCTTCTCTCTGGGCTCGCAGAACCAAGGTTCACAC 683
DB 600 GSCCAAGAACATTCATACCGCATCTCTCTCTCGTAAACACACCGTGGCGGGCC 659
QY 684 CACCGGTGAGCGGTGAGCGAGAAAGACTCTGAACGAGATCGCGGAGAAAGGGCGCA 743
DB 660 TCTTCTTGTCCAGCACCCGAGATCAAGGCG-----ATCGCGGAGAAAGCGGTG 710
QY 744 CACCTTGTCTCAGGTCTTTATTGCTGGGCTCTGCGCGTGGCTAGCTGCTTCTCCCAA 803
DB 711 CACGCGGCTCAGTCTCTCTCTGCGGCACTGTTGCGGCACTCGGTTATCCCA 770
QY 804 GAGCTCCAAACCCCAAGCGCATTTAGTCCAACTTCAAGAGCATTTGAGTCTCCGATGCCGA 863
DB 771 GTCGGTCAACCTTCCGCAATTGGCGAGAACTTCAAGCAGGTCTGGCTCTCGCAGGAG 830
QY 864 CTTTGAAGCATCAATGCCGTTGCCAAGGGTCTGCTTCCGTTTCGTCACCAT 917
DB 831 CGTCGATGCGCTCAGCAAGCTCGGCGAGGGTTCGGGCGCGAGCGCTACACAT 884

RESULT 11
US-10-040-416-5
; Sequence 5, Application US/10040416
; Publication No. US20020160480A1
; GENERAL INFORMATION:
; APPLICANT: OOKURA, TETUYA
; APPLICANT: KASUMI, TAKAFUMI
; APPLICANT: ASABA, Eiji
; TITLE OF INVENTION: ERYTHROSE REDUCTASE, ITS CDNA AND CELL WHICH THE CDNA EXPRESS
; FILE REFERENCE: 217969USOXCIIP
; CURRENT APPLICATION NUMBER: US/10/040,416
; CURRENT FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: US 09/800,487
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: JP 2001-001294
; PRIOR FILING DATE: 2001-01-09
; NUMBER OF SEQ ID NOS: 10

```

; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (717)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (723)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (728)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (734)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-10-004-115A-19

```

Query Match	42.1%	Score 412;	DB 15;	Length 743;
Best Local Similarity	100.0%;	Pred. No. 9.9e-122;		
Matches 412;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	567	CAACAGATCGAGATTCAACCCCTTCGTCGCCAACAGAGGAGCTGGTSCAGTACTGCTGCTTC	625	
Db	1	CAACAGATCGAGATTCAACCCCTTCGTCGCCAACAGAGGAGCTGGTSCAGTACTGCTGCTTC	60	
Qy	627	CAAGAACATTATGCCCGTGGCTACTCTCTCTGGCTCGCAACCAAGGTTCCACCAC	686	
Db	61	CAAGAACATTATGCCCGTGGCTACTCTCTCTGGCTCGCAACCAAGGTTCCACCAC	120	
Qy	687	CGGTAGCGGGTCAGCGAACAAGACTCTGAACGAGATCGCGAGAGGGCGGCAACAC	745	
Db	121	CGGTAGCGGGTCAGCGAACAAGACTCTGAACGAGATCGCGAGAGGGCGGCAACAC	180	
Qy	747	CTTTGCTCAGGTTCTTATTGCTGGGCTCTGCGCGTGGCTACGTCGTTCTCCCAAG	806	
Db	181	CTTTGCTCAGGTTCTTATTGCTGGGCTCTGCGCGTGGCTACGTCGTTCTCCCAAG	240	
Qy	807	CTCCAAACCCCAAGGCATTGAGTCCAACTTCAAGAGCATGAGCTCTCGATGCCGATT	866	
Db	241	CTCCAAACCCCAAGGCATTGAGTCCAACTTCAAGAGCATGAGCTCTCGATGCCGATT	300	
Qy	867	TGAAGCCATCAATGCCGTTGCCAAGGTCGTCACTTCGTTTCGTCACATCAAGAGTAC	926	
Db	301	TGAAGCCATCAATGCCGTTGCCAAGGTCGTCACTTCGTTTCGTCACATCAAGAGTAC	360	
Qy	927	TTTCGGATATGATGTCCTGGCCGAGGAGACGGCCAAAGAACCTGTCTGCGTGA	978	
Db	361	TTTCGGATATGATGTCCTGGCCGAGGAGACGGCCAAAGAACCTGTCTGCGTGA	412	

RESULT 8
US-10-369-493-27083
; Sequence 27083, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 27083
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Neurospora crassa

Query Match	41.7%	Score 408.2	DB 16	Length 765
Best Local Similarity	71.8%	Pred. No. 1.7e-120		
Matches 549	Conservative 0	Mismatches 213	Indels 3	
Qy	46	CCTGGCCTCGCCTTGGTACCTTCGCTAGTCGAAGGTTCCAAAGGGCGAGACCTTAT		
Db	1	CCCGCGTCGGATTTCGACACTTTTGCACACGGGTGCCAAGGGTCCAAAGGGCGAGACCTTA		
Qy	106	GTACCACTGCCCTGAAGACCGGTTACCGTCACCTTGGACTGTGCTCGTGTACTATTA		
Db	61	GTCAAGAAGGCCCTCGAGGTTGGCTACCGTCACCTCGACTGCGCGTGGTTCCTAT		
Qy	166	GAGGCTCAGGTTTGGTGAAGGTTATCCGTGCACTTCTTGAAGGAGAAACCCCTCGGT		
Db	121	GAGGACAGGTCGSCCAGGGCGCTCGCGAGTTCCTTGAGAACCAAGGACGCTG		
Qy	226	GAGGACATCTTCGTCTGCACCAAGGTTGGAAACCACTCCACCCTTATGAGAGAA		
Db	181	GAGGACATCTTCATCTGCACCAAGGTTGGAAACCACTCGACAGAGCCCGAGGAGA		
Qy	286	TGTTCCATTGACGACTCCCTGAAGCGTCTTGACCTTGACTACGTGTGATATGTTT		
Db	241	TGGAGCTTGAGAACTCGCTCGACAAAGCTCAAGGTTGACTACGTGCACTCTT		
Qy	346	CACTGGGCCATTGTCGCGAGAG---AATGGCCAGGTTGAGCCCAAGATTGG		
Db	301	CACTGGGCCATTGTCGCGCGAGAGGAGCGAGGCCACCAACATGCCCAAGATCGG		
Qy	403	GGCAAAATACGTCTATCTCAAGGACTCGACCGAGAACCCCGAGGCCACATGGCGG		
Db	361	GGCAAGTACATCATCAAGAAGGAGCTCACCCGAGAACCCCGAGGCCAACTGGCGG		
Qy	463	GAGNAGATTATGAGGATCGCAAGGCCAGGTCCATTGGTGCTCTCAAATCGAGAC		
Db	421	GAGGATCTCGTGGATGCGGCAAGCGGGTCCATTGGTGTTTCCAACTGGAGAC		
Qy	523	GACCTTCAGAAGATGTCAAAGTTCGCCAAAGTTCATGCCCTACGCCAAACCAATAT		
Db	481	GGCCTGCAGAAGCTCCTCAAGTTTGCTCGCATCAAGCCGACCGTCAAGCCAGAT		
Qy	583	CACCCCTTCGTCGCCAAGGAGCTGGTGAGTACTCTTCTCTCCAAAGACAT		
Db	541	CACCCCTTCCTCCCAACACCGAGCTCGTCGAGTTCTGCTTCAAGAACCAAT		
Qy	643	GTGGCCTACTCTCTCTGGGCTCGCAAGACCAAGGTTCCCAACACGGGTGAGCGG		
Db	601	ACTGCCTACTCTCTCTCGGTCGCAAGACCAAGGTTCCCATCGACGGGCGAGCGG		
Qy	703	GAGNACAAGACTCTGAACGAGATCGCCGAGAAAGGCGGCAACACCCCTTGCTCA		
Db	661	GATGACCCCACTCTCAAGGCCGTTGCTGAGCGCAGCGACACAACTTGCCCA		
Qy	763	ATTGCCCTGGGCTCTGCGCGTGGGTAGTCGTTCTCCCAAGAGC 807		
Db	721	CTGCGCTGGGCTCTCCGACAGAGGTACGTCGTTCTCCCAAGAGC 765		

RESULT 9
US-10-004-115A-18
Sequence 18, Application US/10004115A
Publication No. US20030134402A1
GENERAL INFORMATION:
APPLICANT: ASAKO, HIROYUKI
APPLICANT: MATSUMURA, KENJI
APPLICANT: SHIMIZU, MASATOSHI
APPLICANT: ITO, NOBUYA
APPLICANT: WAKITA, KYUHEI
TITLE OF INVENTION: PROCESS FOR PRODUCING OPTICALLY ACTIVE
TITLE OF INVENTION: 4-HALO-3-HYDROXYBUTANOATE
FILE REFERENCE: 7372-72249
CURRENT APPLICATION NUMBER: US/10/004,115A

Qy 605 AGCTGGTGCAGTACTGTTCTCCAGAACATATATCCCGTGGCCTACTCTCCTCTGG 661
Db 361 AGCTGGTGCAGTACTGTTCTCCAGAACATATATCCCGTGGCCTACTCTCCTCTGG 417

RESULT 7

US-10-004-115A-19
; Sequence 19, Application US/10004115A
; Publication No. US20030134402A1
; GENERAL INFORMATION:
; APPLICANT: ASAKO, HIROYUKI
; APPLICANT: MATSUMURA, KENJI
; APPLICANT: SHIMIZU, MASATOSHI
; APPLICANT: ITO, NOBUYA
; APPLICANT: WAKITA, RYUHEI
; TITLE OF INVENTION: PROCESS FOR PRODUCING OPTICALLY ACTIVE
; FILE REFERENCE: 7372-72249
; CURRENT APPLICATION NUMBER: US/10/004,115A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: JP 2000-372704
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: JP 2001-006144
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: JP 2001-026594
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: JP 2001-175175
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 743
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (434)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (440)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (443)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (448)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (458)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (467)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (475)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (506)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (510)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (517)
; OTHER INFORMATION: a, c, t, g, other or unknown
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (522)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (533)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (535)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (583)..(584)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (607)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (616)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (634)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (639)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (648)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (651)..(652)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (659)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (662)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (664)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (670)..(671)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (690)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (693)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (695)
; OTHER INFORMATION: a, c, t, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (702)
; OTHER INFORMATION: a, c, t, g, other or unknown

FEATURE:
NAME/KEY: modified_base
LOCATION: (463)
OTHER INFORMATION: a, c, t, g, other or unknown
FEATURE:
NAME/KEY: modified_base
LOCATION: (485)
OTHER INFORMATION: a, c, t, g, other or unknown
FEATURE:
NAME/KEY: modified_base
LOCATION: (489)
OTHER INFORMATION: a, c, t, g, other or unknown
FEATURE:
NAME/KEY: modified_base
LOCATION: (491)
OTHER INFORMATION: a, c, t, g, other or unknown
FEATURE:
NAME/KEY: modified_base
LOCATION: (520)
OTHER INFORMATION: a, c, t, g, other or unknown
FEATURE:
NAME/KEY: modified_base
LOCATION: (524)
OTHER INFORMATION: a, c, t, g, other or unknown
FEATURE:
NAME/KEY: modified_base
LOCATION: (593)
OTHER INFORMATION: a, c, t, g, other or unknown
FEATURE:
NAME/KEY: modified_base
LOCATION: (602)
OTHER INFORMATION: a, c, t, g, other or unknown
FEATURE:
NAME/KEY: modified_base
LOCATION: (682)
OTHER INFORMATION: a, c, t, g, other or unknown
US-10-004-115A-15

Query Match 60.2%; Score 588.6; DB 15; Length 697;
Best Local Similarity 94.4%; Pred. No. 1.5e-178;
Matches 611; Conservative 0; Mismatches 35; Indels 1; Gaps 1;
QY 4 TCTAACGGAAGACTTTCACATTCAGCAACGGCGTCAAGATTCCTGGCGTCGGCTTGGT 63
DB 52 TCCAAACGGAANACTTTCACACTGAGCAACGGCGTCAAAATTCCTGGCGTCGGCTTGGT 111
QY 64 ACCTTCGGTACTGGAAGGTTCCAAAGGGGAGACCTATATCTGTGTCAACACTGCCCTCAAG 123
DB 112 ACCTCGGTACTGGAAGGTTCCAAAGGGGAGACCTATATCTGTGTCAACACTGCCCTCAAG 171
QY 124 ACCGGTTACCGTCACTTGGACTGTGCTGGTACTACCTGACGAGGCTGAGTGGTGAG 183
DB 172 ACCGGTTACCGTCACTTGGACTGTGCTGGTACTACCTGACGAGGCTGAGTGGTGAG 231
QY 184 GGTATCCGTGACTTCCCTGAAGGAGAACCCCTCGGTGAAGGCTGAGGACATCTTCGTCTGC 243
DB 232 GGTATCCGTGACTTCCCTGAAGGAGAAACCCCTCGGTGAAGGCTGAGGACATCTTCGTCTGC 291
QY 244 ACCAAGGTGTGGAACCAACCTCCACCGTTATGAGGACGTTCCTGGTCCATTGACGATCC 303
DB 292 ACCAAGGTGTGGAACCAACCTCCACCGTTATGAGGACGTTCCTGGTCCATTGACGATCC 351
QY 304 CTGAAGCGTCTTGACCTGACTGACTGTTGATATGTTCTCGTTCACCTGGCCCATGCTGCC 363
DB 352 CTGAAGCGTCTTGACCTGACTGACTGTTGATATGTTCTCGTTCACCTGGCCCATGCTGCC 411
QY 364 GAGAAGATGCGCAGGCTGAGCCCAAGATTGGCCCTCGACCGCAATACGTCATCTCAAG 423
DB 412 GAAAAAATGCGCAGGCTGAGCCCAAAATTTGGCCCTGACGCAAAATACNTCTCAAG 471
QY 424 GACCTGACCGAGACCCCGACCCACATGCGCGCTATGAGAGATTTATGAGATCCG 483
DB 472 GACCTGACCGA-AAANCCCNANCCCACTGGCGCTATGGAAAAATTTTNTGANGATCCC 530

QY 484 AAGCCAGGTCCATTGGTGTCTCCAACTGGACCATTCGCCACCTTGAGAGATGTCCAG 543
DB 531 AAGCCAGGTCCATTGGTGTCTCCAACTGGACCATTCGCCACCTTGAGAGATGTCCAG 590
QY 544 TTGCGCAAGGTTCATGCTTCAGCCCAACAGATCGAGATTCACCCCTTCTGCGCCCAACGAG 603
DB 591 TTGCGCAAGGTTCATGCTTCAGCCCAACAGATCGAGATTCACCCCTTCTGCGCCCAACGAG 650
QY 604 GAGCTGGTGCAGTACTGCTTCTCCAGAACATTTATGCCCGTGGCCTA 650
DB 651 GAGCTGGTGCAGTACTGCTTCTCCAGAACATTTATGCCCGTGGCCTA 697

RESULT 6

US-10-004-115A-22
Sequence 22, Application US/10004115A
Publication No. US20030134402A1
GENERAL INFORMATION:
APPLICANT: ASAKO, HIROYUKI
APPLICANT: MATSUMURA, KENJI
APPLICANT: SHIMIZU, MASATOSHI
APPLICANT: ITO, NOBUYA
APPLICANT: WAKITA, RYUHEI
TITLE OF INVENTION: PROCESS FOR PRODUCING OPTICALLY ACTIVE
FILE REFERENCE: 7372-72249
CURRENT APPLICATION NUMBER: US/10/004,115A
CURRENT FILING DATE: 2002-10-23
PRIOR APPLICATION NUMBER: JP 2000-372704
PRIOR FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: JP 2001-006144
PRIOR FILING DATE: 2001-01-15
PRIOR APPLICATION NUMBER: JP 2001-026594
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: JP 2001-175175
PRIOR FILING DATE: 2001-06-11
NUMBER OF SEQ ID NOS: 37
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 22
LENGTH: 417
TYPE: DNA
ORGANISM: Escherichia coli
US-10-004-115A-22

Query Match 42.6%; Score 417; DB 15; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.9e-123;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 CCAAGGTGTGGAACCAACCTCCACCGTTATGAGGACGTCTCTGGTCCATTGACGACTCCC 304
DB 1 CCAAGGTGTGGAACCAACCTCCACCGTTATGAGGACGTCTCTGGTCCATTGACGACTCCC 60
QY 305 TGAAGCGTCTTGGACTTGACTACGTGTTGATATGTTCTCGTTCACCTGGCCCATTTGCTGCCG 364
DB 61 TGAAGCGTCTTGGACTTGACTACGTGTTGATATGTTCTCGTTCACCTGGCCCATTTGCTGCCG 120
QY 365 AGAAGATGCGCAGGTTGAGCCCAAGATTCGCTGAGGCAATACGTCATTTCTCAAGG 424
DB 121 AGAAGATGCGCAGGTTGAGCCCAAGATTTGGCCCTGAGCGCAAAATACGTCATTTCTCAAGG 180
QY 425 ACCTGACCGAGAACCCCGAGCCACATGCGCGCTATGAGAGAGATTTATGAGGATCGCA 484
DB 181 ACCTGACCGAGAACCCCGAGCCACATGCGCGCTATGAGAGAGATTTATGAGGATCGCA 240
QY 485 AGCCAGGTCATTTGGTGTCTCCAACTGGAACATTTGCCGACCTTGAGAGATGTCCAGT 544
DB 241 AGCCAGGTCATTTGGTGTCTCCAACTGGAACATTTGCCGACCTTGAGAGATGTCCAGT 300
QY 545 TCGCCAAAGGTTCATGCTCAGCCCAACAGATCGAGATTCACCCCTTCTGCGCCCAACGAGG 604
DB 301 TCGCCAAAGGTTCATGCTCAGCCCAACAGATCGAGATTCACCCCTTCTGCGCCCAACGAGG 360

Db 601 GAGGAGCTGGTGAAGTCTCTTCTCAAGAACATTTATGCCCCGTGCTACTCTCTCTG 660
Qy 661 GGCTCGAGAACCAAGGTTCCCAACACCGGTGAGCGGGTCAGCGAGAACAAAGACTCTGAAC 720
Db 661 GGCTCGAGAACCAAGGTTCCCAACACCGGTGAGCGGGTCAGCGAGAACAAAGACTCTGAAC 720
Qy 721 GAGATCGCCGAGAGGGGCGGACACCCCTTCTCAGGTTCTTATTCGCTGGGCTCTGCGC 780
Db 721 GAGATCGCCGAGAGGGGCGGACACCCCTTCTCAGGTTCTTATTCGCTGGGCTCTGCGC 780
Qy 781 CGTGGCTACGTCGTTCTCCCAAGAGCTCCAAACCCCAAGCGCATTCAGTCCCACTTCAAG 840
Db 781 CGTGGCTACGTCGTTCTCCCAAGAGCTCCAAACCCCAAGCGCATTCAGTCCCACTTCAAG 840
Qy 841 AGCATTTAGGCTCTCCGATGCGGACTTTGAAGCATTCAATCCGTTGCGAAGGTCGTCAC 900
Db 841 AGCATTTAGGCTCTCCGATGCGGACTTTGAAGCATTCAATCCGTTGCGAAGGTCGTCAC 900
Qy 901 TTCGTTTTCGTAACATGAAGGATCTTTCCGATATGATGTCGCGCCGAGGAGACCGCC 960
Db 901 TTCGTTTTCGTAACATGAAGGATCTTTCCGATATGATGTCGCGCCGAGGAGACCGCC 960
Qy 961 AAGAACCTGTCGCTGA 978
Db 961 AAGAACCTGTCGCTGA 978

RESULT 3
US-10-608-533-2
; Sequence 2, Application US/10608533
; Publication No. US20040091902A1
; GENERAL INFORMATION:
; APPLICANT: ASAKO, HIROYUKI
; APPLICANT: SHIMIZU, MASATOSHI
; TITLE OF INVENTION: MODIFIED REDUCTASE AND ITS GENE
; FILE REFERENCE: Q76265
; CURRENT APPLICATION NUMBER: US/10/608,533
; PRIOR FILING DATE: 2003-06-30
; PRIOR APPLICATION NUMBER: JP 2002-193074
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 978
; TYPE: DNA
; ORGANISM: Penicillium citrinum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(975)
US-10-608-533-2

Query Match 100.0%; Score 978; DB 17; Length 978;
Best Local Similarity 100.0%; Pred. No. 1e-303;
Matches 978; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCTTAACGGAAGACATTTACATTGAGCAACGGCGTCAAGATTCTGCGCTCGGCTTT 60
Db 1 ATGCTTAACGGAAGACATTTACATTGAGCAACGGCGTCAAGATTCTGCGCTCGGCTTT 60
Qy 61 GGTACCTTCGCTAGTGAAGTTCCAGGGCGAGACCTATATCTGCTGTCACCACTGCGCTG 120
Db 61 GGTACCTTCGCTAGTGAAGTTCCAGGGCGAGACCTATATCTGCTGTCACCACTGCGCTG 120
Qy 121 AAGACCGGTTACCGTCACTTGGACTGTCCTGGTATCTACTGACGAGGAGGTTGGT 180
Db 121 AAGACCGGTTACCGTCACTTGGACTGTCCTGGTATCTACTGACGAGGAGGTTGGT 180
Qy 181 GAGGGTATCCGTGACTTCTCGTGAAGGAAACCCCTCGGTGAAGCGTGAAGGACATCTTCGTC 240
Db 181 GAGGGTATCCGTGACTTCTCGTGAAGGAAACCCCTCGGTGAAGCGTGAAGGACATCTTCGTC 240
Qy 241 TGCACCAAGGTGGAACCAACCTCCACCGTTATGAGGACGCTCTCTGTCATTCAGCAC 300

Db 241 TGCACCAAGGTGGAACCAACCTCCACCGTTATGAGGAGGTCCTCTGTCATTCAGCAC 300
Qy 301 TCCCTGAAGCGCTTTGGACTTGAATGATGTTCTCTGTTCACTGCCCCATTTGCT 360
Db 301 TCCCTGAAGCGCTTTGGACTTGAATGATGTTCTCTGTTCACTGCCCCATTTGCT 360
Qy 361 GCCGAGAAAGATGGCCAGGGTGAGCCCCAAGATTGGCCCTGACGGCAATACGTCATCTC 420
Db 361 GCCGAGAAAGATGGCCAGGGTGAGCCCCAAGATTGGCCCTGACGGCAATACGTCATCTC 420
Qy 421 AAGGACCTTGAACCGAGAACCCCGAGCCACATGCGCGCGCTATGGAAGAGATTATGAGAT 480
Db 421 AAGGACCTTGAACCGAGAACCCCGAGCCACATGCGCGCGCTATGGAAGAGATTATGAGAT 480
Qy 481 CGCAGGCCAGGTCCTATGTTGTTCTCCAACTGGACCATTTGCCGACCTTGAGAAGATGCC 540
Db 481 CGCAGGCCAGGTCCTATGTTGTTCTCCAACTGGACCATTTGCCGACCTTGAGAAGATGCC 540
Qy 541 AAGTTTCGCAAGGTCATGCTTCAGCCCAACAGATCGAGATTCAACCCCTTCTGCCCCAAC 600
Db 541 AAGTTTCGCAAGGTCATGCTTCAGCCCAACAGATCGAGATTCAACCCCTTCTGCCCCAAC 600
Qy 601 GAGGAGCTGGTGCAGTACTGCTTCTCCAGAACATTTATGCGCGTGGCTACTCTCTCTG 660
Db 601 GAGGAGCTGGTGCAGTACTGCTTCTCCAGAACATTTATGCGCGTGGCTACTCTCTCTG 660
Qy 661 GGCTCGAGAACCAAGGTTCCCAACACCGGTGAGCGGGTCAGCGAGAACAAAGACTCTGAAC 720
Db 661 GGCTCGAGAACCAAGGTTCCCAACACCGGTGAGCGGGTCAGCGAGAACAAAGACTCTGAAC 720
Qy 721 GAGATCGCGGAGAGGGGCGGCAACACCTTCTCAGGTTCTTATGCTGGGCTCTGCGC 780
Db 721 GAGATCGCGGAGAGGGGCGGCAACACCTTCTCAGGTTCTTATGCTGGGCTCTGCGC 780
Qy 781 CGTGGCTAGTCTGTTCTCCCAAGAGTCCCAACCGGAGCGCATTTAGTCCCACTTCAAG 840
Db 781 CGTGGCTAGTCTGTTCTCCCAAGAGTCCCAACCGGAGCGCATTTAGTCCCACTTCAAG 840
Qy 841 AGCATTTAGCTCTCCGATGCGGACTTTGAAGCCATTAATGCGTTGCCAAGGGTCTGAC 900
Db 841 AGCATTTAGCTCTCCGATGCGGACTTTGAAGCCATTAATGCGTTGCCAAGGGTCTGAC 900
Qy 901 TTCGTTTTCGTAACATGAAGGATCTTTCCGATATGATGTCGCGCCGAGGAGACCGCC 960
Db 901 TTCGTTTTCGTAACATGAAGGATCTTTCCGATATGATGTCGCGCCGAGGAGACCGCC 960
Qy 961 AAGAACCTGTCGCTGA 978
Db 961 AAGAACCTGTCGCTGA 978

RESULT 4
US-10-004-115A-28
; Sequence 28, Application US/10004115A
; Publication No. US20030134402A1
; GENERAL INFORMATION:
; APPLICANT: ASAKO, HIROYUKI
; APPLICANT: MATSUMURA, KENJI
; APPLICANT: SHIMIZU, MASATOSHI
; APPLICANT: ITO, NOBUYA
; APPLICANT: WAKITA, RYUHEI
; TITLE OF INVENTION: PROCESS FOR PRODUCING OPTICALLY ACTIVE
; FILE REFERENCE: 7372-72249
; CURRENT APPLICATION NUMBER: US/10/004,115A
; PRIOR FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: JP 2000-372704
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: JP 2001-006144
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: JP 2001-026594
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: JP 2001-175175

QY 1 ATGTCTAACGGAAGACTTTTCAATTGAGCAACGGCGTCAAGATTCTTGGCGTCGGCTTT 60
Db |||||
1 ATGTCTAACGGAAGACTTTTCAATTGAGCAACGGCGTCAAGATTCTTGGCGTCGGCTTT 60
QY 61 GGTACCTTCGCTAGTGAAGTTTCAAGGGCGAGACCTTACTCTGTGCACCACTTGCCTG 120
Db |||||
61 GGTACCTTCGCTAGTGAAGTTTCAAGGGCGAGACCTTACTCTGTGCACCACTTGCCTG 120
QY 121 AAGACCGGTTACCGTCACTTGGACTGTGCTGTGCTTACCTGAAACGAGGCTGAGGTGGT 180
Db |||||
121 AAGACCGGTTACCGTCACTTGGACTGTGCTGTGCTTACCTGAAACGAGGCTGAGGTGGT 180
QY 181 GAGGATTCGGTGAAGTTTCTGAGGAGAACCCCTCGGTGAAGCTGAGGACATCTTCGTC 240
Db |||||
181 GAGGATTCGGTGAAGTTTCTGAGGAGAACCCCTCGGTGAAGCTGAGGACATCTTCGTC 240
QY 241 TGCACCAAGGTGTGGAACCACTTCCACCGTTATGAGGACGTCCTCTGCTGCTTGCACG 300
Db |||||
241 TGCACCAAGGTGTGGAACCACTTCCACCGTTATGAGGACGTCCTCTGCTGCTTGCACG 300
QY 301 TCCCTGAAGCGTCTTGACCTGACTAGTGTGATATGCTTCTGCTGCTGCTGCTGCTGCT 360
Db |||||
301 TCCCTGAAGCGTCTTGACCTGACTAGTGTGATATGCTTCTGCTGCTGCTGCTGCTGCT 360
QY 361 GCCGAGAAAGTATGCGGAGGTGAGCCCAAGATTGCGCTGACCGCAAAATACGTCATTCTC 420
Db |||||
361 GCCGAGAAAGTATGCGGAGGTGAGCCCAAGATTGCGCTGACCGCAAAATACGTCATTCTC 420
QY 421 AAGGACCTGACCGAGAACCCCGAGCCACATGCGCGCTATGCGGCAAGATTATGAGGAT 480
Db |||||
421 AAGGACCTGACCGAGAACCCCGAGCCACATGCGCGCTATGCGGCAAGATTATGAGGAT 480
QY 481 CGCAAGGCGAGTCCATTGGTGTCTCAACTGGAACCATGCGGCACTTGAAGAGATGCTCC 540
Db |||||
481 CGCAAGGCGAGTCCATTGGTGTCTCAACTGGAACCATGCGGCACTTGAAGAGATGCTCC 540
QY 541 AAGTTCGCAAGGTCAATGCTTCCACCAACGAGATCGAGATTCAACCGCTTCTGCGCCAAC 600
Db |||||
541 AAGTTCGCAAGGTCAATGCTTCCACCAACGAGATCGAGATTCAACCGCTTCTGCGCCAAC 600
QY 601 GAGGAGCTGTGTCAGTACTGCTTCTCAAGAACATATGCGGTCGCTGCTGCTGCTGCTG 660
Db |||||
601 GAGGAGCTGTGTCAGTACTGCTTCTCAAGAACATATGCGGTCGCTGCTGCTGCTGCTG 660
QY 661 GGCTCGCAGAACAGGTTTCCACCAACCGGTGAGCGGCTCAGCGAGAACAGACTCTGAAC 720
Db |||||
661 GGCTCGCAGAACAGGTTTCCACCAACCGGTGAGCGGCTCAGCGAGAACAGACTCTGAAC 720
QY 721 GAGATCGCGAGAGAGGGGGCAACACCTTGTCTGCTGAGTTCTTATGCTGGGCTGCTGCG 780
Db |||||
721 GAGATCGCGAGAGAGGGGGCAACACCTTGTCTGCTGAGTTCTTATGCTGGGCTGCTGCG 780
QY 781 CGTGGCTACGTCGTTCTCCCAAGAGCTTCCAAACCGCTGAGCGCATTTGAGTCCAACTTCAAG 840
Db |||||
781 CGTGGCTACGTCGTTCTCCCAAGAGCTTCCAAACCGCTGAGCGCATTTGAGTCCAACTTCAAG 840
QY 841 AGCAATTGAGCTCTCGGATGCGCATTTTGAAGGCAATCAATGCGGTTGCCAAGGTCGTCAC 900
Db |||||
841 AGCAATTGAGCTCTCGGATGCGCATTTTGAAGGCAATCAATGCGGTTGCCAAGGTCGTCAC 900
QY 901 TTCCGTTTCGTCACATGAAGATACCTTTCGATATGATGTCTGCGCCGAGGAGACCGCC 960
Db |||||
901 TTCCGTTTCGTCACATGAAGATACCTTTCGATATGATGTCTGCGCCGAGGAGACCGCC 960
QY 961 AAGAACCTGTCTGGGTGA 978
Db |||||
961 AAGAACCTGTCTGGGTGA 978

RESULT 2

US-10-327-108-4

; Sequence 4, Application US/10327:108

Publication No. US20030186400A1
GENERAL INFORMATION:
APPLICANT: ASAKO, HIROYUKI
APPLICANT: WAKITA, RYUHEI
APPLICANT: ITOH, NOBUVA
TITLE OF INVENTION: METHOD FOR PRODUCING OPTICALLY ACTIVE
TITLE OF INVENTION: 2-HYDROXYCYCLOALKANECARBOXYLIC ACID ESTER
FILE REFERENCE: Q73517
CURRENT APPLICATION NUMBER: US/10/327,108
CURRENT FILING DATE: 2003-12-24
PRIOR APPLICATION NUMBER: JP 2001-395884
PRIOR FILING DATE: 2001-12-27
PRIOR APPLICATION NUMBER: JP 2001-395885
PRIOR FILING DATE: 2001-12-27
PRIOR APPLICATION NUMBER: JP 2002-107648
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 978
TYPE: DNA
ORGANISM: Penicillium citrinum
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(978)
US-10-327-108-4

Query Match 100.0%; Score 978; DB 15; Length 978;

Best Local Similarity 100.0%; Pred. No. 1e-303;

Matches 978; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCTAACGGAAGACTTTTCAATTGAGCAACGGCGTCAAGATTCTTGGCGTCGGCTTT 60
Db |||||
1 ATGTCTAACGGAAGACTTTTCAATTGAGCAACGGCGTCAAGATTCTTGGCGTCGGCTTT 60
QY 61 GGTACCTTCGCTAGTGAAGTTTCAAGGGCGAGACCTTACTCTGTGCACCACTTGCCTG 120
Db |||||
61 GGTACCTTCGCTAGTGAAGTTTCAAGGGCGAGACCTTACTCTGTGCACCACTTGCCTG 120
QY 121 AAGACCGGTTACCGTCACTTGGACTGTGCTGTGCTTACCTGAAACGAGGCTGAGGTGGT 180
Db |||||
121 AAGACCGGTTACCGTCACTTGGACTGTGCTGTGCTTACCTGAAACGAGGCTGAGGTGGT 180
QY 181 GAGGATTCGGTGAAGTTTCTGAGGAGAACCCCTCGGTGAAGCTGAGGACATCTTCGTC 240
Db |||||
181 GAGGATTCGGTGAAGTTTCTGAGGAGAACCCCTCGGTGAAGCTGAGGACATCTTCGTC 240
QY 241 TGCACCAAGGTGTGGAACCACTTCCACCGTTATGAGGACGTCCTCTGCTGCTTGCACG 300
Db |||||
241 TGCACCAAGGTGTGGAACCACTTCCACCGTTATGAGGACGTCCTCTGCTGCTTGCACG 300
QY 301 TCCCTGAAGCGTCTTGACCTGACTAGTGTGATATGCTTCTGCTGCTGCTGCTGCTGCT 360
Db |||||
301 TCCCTGAAGCGTCTTGACCTGACTAGTGTGATATGCTTCTGCTGCTGCTGCTGCTGCT 360
QY 361 GCGGAGAAAGTGGCCAGGCTGAGCCCAAGATTGCGCTGAGCGCAAAATACGTCATTCTC 420
Db |||||
361 GCGGAGAAAGTGGCCAGGCTGAGCCCAAGATTGCGCTGAGCGCAAAATACGTCATTCTC 420
QY 421 AAGGACCTGACCGAGAACCCCGAGCCACATGCGCGCTATGCGGCAAGATTATGAGGAT 480
Db |||||
421 AAGGACCTGACCGAGAACCCCGAGCCACATGCGCGCTATGCGGCAAGATTATGAGGAT 480
QY 481 CGCAAGGCGAGTCCATTGGTGTCTCAACTGGAACCATGCGGCACTTGAAGAGATGCTCC 540
Db |||||
481 CGCAAGGCGAGTCCATTGGTGTCTCAACTGGAACCATGCGGCACTTGAAGAGATGCTCC 540
QY 541 AAGTTCGCAAGGTCAATGCTTCCACCAACGAGATCGAGATTCAACCGCTTCTGCGCCAAC 600
Db |||||
541 AAGTTCGCAAGGTCAATGCTTCCACCAACGAGATCGAGATTCAACCGCTTCTGCGCCAAC 600
QY 601 GAGGAGCTGTGTCAGTACTGCTTCTCAAGAACATATGCGGTCGCTGCTGCTGCTGCTG 660
Db |||||

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: May 29, 2004, 11:34:38 ; Search time 495 Seconds
(without alignments)
8987.526 Million cell updates/sec

Title: US-10-004-115B-2

Perfect score: 978

Sequence: 1 atgtctaacgaaagacttt.....ccaagaactgtctgcgtga 978

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2860401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq:
14: /cgn2_6/ptodata/1/pubna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/1/pubna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/1/pubna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	978	100.0	978	15	US-10-004-115A-2
2	978	100.0	978	15	Sequence 2, Appli
3	978	100.0	978	17	Sequence 4, Appli
4	978	100.0	996	15	Sequence 2, Appli
5	588.6	60.2	697	15	US-10-004-115A-28
6	417	42.6	417	15	Sequence 15, Appl
7	412	42.1	743	15	US-10-004-115A-22
8	408.2	41.7	765	16	Sequence 19, Appl
9	265	27.1	331	15	US-10-004-115A-19
10	199.6	20.4	972	13	Sequence 27083, A
11	144.2	14.7	1121	14	Sequence 18, Appl
12	142.6	14.6	1077	14	US-09-734-237B-31
13	130.8	13.4	1289	13	Sequence 31, Appl
14	129.4	13.2	1289	13	US-10-040-416-5
					Sequence 5, Appli
					Sequence 3, Appli
					Sequence 17144, A
					Sequence 25069, A

15 129 13.2 1185 13 US-10-425-114-30950
16 129 13.2 1291 13 US-10-425-114-20425
17 119.2 12.2 1335 17 US-10-641-643-1010
18 119.2 12.2 1367 9 US-09-864-864-332
19 119.2 12.2 1367 13 US-10-342-887-641
20 119.2 12.2 1367 13 US-10-172-118-641
21 119.2 12.2 1367 15 US-10-171-581-1
22 113.2 11.6 942 13 US-08-734-237B-43
23 113 11.6 1337 13 US-10-205-331-29
24 113 11.6 1339 12 US-10-152-319A-1492
25 113 11.6 1339 16 US-10-191-803-12
26 111.8 11.4 1313 9 US-09-925-301-47
27 111 11.3 972 15 US-10-321-204-11
28 111 11.3 972 15 US-10-321-204-13
29 111 11.3 972 15 US-10-321-204-16
30 111 11.3 1290 15 US-10-321-204-8
31 111 11.3 1290 15 US-10-198-846-12927
32 110.8 11.3 1171 13 US-10-425-114-5041
33 110.8 11.3 1226 13 US-10-425-114-15182
34 109.4 11.2 1064 15 US-10-321-204-12
35 109.4 11.2 1194 15 US-10-321-204-15
36 109.4 11.2 1219 13 US-10-403-161-85
37 109.4 11.2 1219 15 US-10-321-204-10
38 109.4 11.2 1225 13 US-10-403-161-91
39 109.4 11.2 1596 13 US-10-116-802-150
40 108.6 11.1 1006 13 US-10-114-270-147
41 108.2 11.1 972 13 US-09-734-237B-33
42 106.6 10.9 1119 9 US-09-850-716A-434
43 106.6 10.9 1119 9 US-09-897-778-434
44 106.6 10.9 1119 13 US-10-007-700-434
45 106.6 10.9 1119 15 US-10-117-982-434

ALIGNMENTS

RESULT 1

US-10-004-115A-2
; Sequence 2, Application US/10004115A
; Publication No. US20030134402A1
; GENERAL INFORMATION:
; APPLICANT: ASAKO, HIROYUKI
; APPLICANT: MATSUMURA, KENJI
; APPLICANT: SHIMIZU, MASATOSHI
; APPLICANT: ITO, NOBUYA
; APPLICANT: WAKITA, RYUHEI
; TITLE OF INVENTION: PROCESS FOR PRODUCING OPTICALLY ACTIVE
; FILE REFERENCE: 4-HALO-3-HYDROXYBUTANOATE
; CURRENT APPLICATION NUMBER: US/10/004,115A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: JP 2000-372704
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: JP 2001-006144
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: JP 2001-026594
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: JP 2001-175175
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 978
; TYPE: DNA
; ORGANISM: Penicillium citrinum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(975)
US-10-004-115A-2

Query Match 100.0%; Score 978; DB 15; Length 978;
Best Local Similarity 100.0%; Pred. No. 1e-303;
Matches 978; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

JOURNAL
COMMENT

Unpublished (2002)
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 638 Std Error: 0.00
Plate: 13 row: N column: 5
Seq primer: M3rev.

FEATURES
source

Location/Qualifiers
1..638
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="barke"
/sub_species="vulgare"
/db_xref="GABI:263815"
/db_xref="taxon:112509"
/clone="H113N05"
/tissue_type="endosperm early"
/dev_stage="0-16 hours after imbibition"
/lab_host="XL10-Gold"
/clone_lib="H1"
/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable."

ORIGIN

Query Match 15.2%; Score 148.8; DB 13; Length 638;
Best Local Similarity 55.1%; Pred. No. 3.6e-22;
Matches 335; Conservative 0; Mismatches 267; Indels 6; Gaps 2;

QY 59 TTGTTACCTTCGCTAGTGAAGGTTCCAAAGGCGGAGACCTATACGCTGTCCACCACTGCC 118
DB 15 TTGTTTCGGACATCGCAAGCTAGCTGTGCTGAGTGGCGGCTCGCTCTCGAGGCC 74
QY 119 TGAAGACCGTTACCGTCACTTGGACTGTGCTGCTACTACTCTGAACGAGGGTGAAGTTG 178
DB 75 TCAAGCTGGCTACCGCCACCTTGATTCGCCAAGGCTACGGCAACAGAGAGATCG 134
QY 179 GTGAGGTTACCGTAC---TTCCTGAGGAGAACCCCTCGGTGAACGCTGAGACATCT 235
DB 135 CCCAGGCCCTCAAGAAGGCTTTCGGCGCGAGGTCCTTAACCTCAAGCGTGAGGACGCT 194
QY 236 TCGTCTGCACCAAGGTGTGGAACCACTCCACCGCTTATGAGGACGCTCTCTGTGCCATTG 295
DB 195 TCATCACATCCAAGCTGTGGAACCTGCAGACACCGCCCAAGGACGTGCCGGCGCTCTCG 254
QY 296 AGCACTCCTGAAGCGCTTGGACTGACTAGTGTATGTTCTGCTGTTACCTGAGGCCCA 355
DB 255 AGCACTGCTGGCTGAGTGTGGCTCGACTACCTAGACCTTTACCTGCTTCACTTCCCGC 314
QY 356 TTGCTGCCGCAAGAAGTATGCGCAGGCTGAGCCCAAGATTGGCCCTGAGCGCAAAATACGTC 415
DB 315 TCGCTTTTCGAGGCGCGCGATGTCCACACACCTCTTCCCTCTCGAGGGCGAGAGCG 374
QY 416 TTCTCAAGACCTGACCGAGAACCCGAGCCGCAATGCGCGCTATGAGAGAGATTATG 475
DB 375 TCAAGTCTCGACACGCTCTTATGCTCGACATGGAAGGCCATCACAGCTGCCCA 434
QY 476 AGGATCGCAAGCCAGGTCCTATGTTGCTCCAACTGACCATTTGCCGACCTTCAGAGA 535
DB 435 AGGA---GAAGGCGCGCTGCTGGTGTCTCAACCAACCAAGAGGACCTTCAGGCCA 491
QY 536 TGTCGAAGTTCGCCAAGGTCATGCTCAGCCCAACCAAGATCGAGATTCAACCCCTTCCTGC 595
DB 492 TCATCGAGCGGACCGCGGCTCACTCCCGCTGCAACCAAGATCGAGCGCCACCCCGCTCTCC 551

QY 596 CCAACGAGGAGCTGCTGAGTACTGTTCTTCCAAAGCAATTATGCCCTGGCTACTCTC 655
DB 552 TCACGCCAGAGCTCATCCAGTACTCCAGGAGAGAACATCCACATCCAGGACTCCG 611
QY 656 CTCTGGGC 663
DB 612 CCTTCGGC 619

RESULT 15
LOCUS

AA697426 758 bp mRNA linear EST 19-APR-2001
HL02448.5prime HL Drosophila melanogaster head BlueScript
Drosophila melanogaster cDNA clone HL02448 5prime, mRNA sequence.

ACCESSION

AA697426

VERSION

AA697426.1 GI:2700355

KEYWORDS

EST.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 758)
Harvey D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S. and Rubin G.M.

TITLE

BDGP/HEMI Drosophila EST Project

JOURNAL

Unpublished (2001)

COMMENT

Contact: Stapleton, M.
BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

Plate: 24 row: D column: 12

High quality sequence stop: 630.

FEATURES

source

1..758
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="HL02448"
/sex="male and female"
/dev_stage="adult"
/lab_host="SOLR"
/clone_lib="HL Drosophila melanogaster head BlueScript"
/note="Organ: head-brain & sensory organ; Vector:
BlueScript SK; Site 1: EcoRI; Site 2: XhoI; Constructed
using Stratagene ZAP-cDNA Synthesis kit. Oligo dT-primed
and directionally cloned at EcoRI and XhoI in BlueScript
SK(+/-)"

ORIGIN

Query Match 15.2%; Score 148.6; DB 9; Length 758;
Best Local Similarity 54.8%; Pred. No. 4.4e-22;
Matches 345; Conservative 0; Mismatches 269; Indels 15; Gaps 2;

QY 83 CCAAGGGCGAGACCTATCTGTCACCACTGCCCTGAAGACCGGTTACCGTCACTTGG 142
DB 143 CCAAGGGTCAGGTCACCGGCTGTCAAAGTTGCCATTGATCCCGATACAGGCATTTG 202
QY 143 ACTGTCCCTGGTACTACTTGAACGAGGTTGGTGGAGGTATCCGTGACTTCCCTGA 202
DB 203 ACTGTCCCTATGTATACCAAAACGAGGATGAGTGGAGATGGTGT---TGAGGCCAAGA 259
QY 203 AGGAGAACCCCTCGGTGAACGCTGAGGACATCTTGTCTGACCAAGGTGTGGAACACC 262
DB 260 TCAAGGAGGCGGTGTCAAGCGTGAGGATCTGTTCATCACCAAGCAAACTGTGGAACACT 319
QY 263 TCACCGTTATGAGGACGTCCTCTGTGTCATTGACGACTCCCTGAGAGCGTCTTGACTTG 322
DB 320 TCATCGCCCGGATCTTGTCAAGTCGGCATTTGGAGAACAATTTGAGCTCCCTGGAAGCTGA 379

```
source
1. 189
/organism="Emericella nidulans"
/mol_type="mRNA"
/strain="FGSC A26"
/db_xref="taxon:162425"
/clones="h1h04a1"
/tissue_type="vegetative mycelia, asexual structures"
/clone_lib="Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library"
/note="vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript 3' end of cDNA cloned into XhoI site of pBluescript"

ORIGIN
Query Match 15.6%; Score 152.2; DB 9; Length 189;
Best Local Similarity 87.8%; Pred. No. 3.3e-23;
Matches 166; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 732 GAAGGGGGCAACCCCTTGTCTAGGTTCTTATTGCTGGGCTCTGGCGTGGCTACGT 791
Db 1 GAAGGGGGCAACCCCTTGTCTAGGTTCTTATTGCTGGGCTCTGGCGTGGCTACGT 60
QY 792 CGTTCTCCCCAAGAGCTCCAAACCCCAAGCGGATTCAGTCCAACTTCAAGAGCAATTGAGCT 851
Db 61 CGTCTGCCCAAGAGCTCCAAACCCCGCGCATTCAGTCCAACTTCAAGAGCAATTGAGCT 120
QY 852 CTCGGATGCCGACTTGAAGCCATCAATCGGTTGCCAAGGGTGGTCACTTCGGTTTCGT 911
Db 121 GTCCAGAGGACTACGAGGCGGTCACGCTGTCGCCAAGGGCGTCACTTCGGTTTCGT 180
QY 912 CAACATGAA 920
Db 181 CAACATGAA 189

RESULT 13
AA263299
LOCUS
DEFINITION
Drosophila melanogaster embryo Bluescript
Drosophila melanogaster cDNA clone LD06393 5 similar to CG6084:
FBan006084 GO:[enzyme (GO:0003824)] located on: 3L 68C9-68C9;
08/13/2002, mRNA sequence.
AA263299
AA263299.2 GI:13766074
EST.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
On Jan 24, 1997 this sequence version replaced gi:1899358.
Other ESTs: LD06393.3prime
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003544: arm3L [11370031,11651899]
estimated-cyto:68C9-68C15: 04/10/2001
Plate: LD.63 row: H column: 9
High quality sequence stop: 701.
Location/Qualifiers
1. 779
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="BDGP EST:BDcIn005720"
/db_xref="taxon:7227"

FEATURES
source
1. 189
/organism="Emericella nidulans"
/mol_type="mRNA"
/strain="FGSC A26"
/db_xref="taxon:162425"
/clones="h1h04a1"
/tissue_type="vegetative mycelia, asexual structures"
/clone_lib="Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library"
/note="vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript 3' end of cDNA cloned into XhoI site of pBluescript"

ORIGIN
Query Match 15.2%; Score 149; DB 9; Length 779;
Best Local Similarity 56.0%; Pred. No. 3.6e-22;
Matches 331; Conservative 0; Mismatches 245; Indels 15; Gaps 2;

QY 83 CCAAGGGGAGACCTTACTGCTGTCACACCTGCTCCCTGAAGACCGTTACCGTCACTTGG 142
Db 143 CCAAGGGGTGAGTCCAGGCTGTCAAAGTTGCCATTGATCCCGGATACAGGCCACATTG 202
QY 143 ACTGTGCTGTGTTACTTACCTGACGAGGGTGTAGTTTGGTGGGTATCCGTGACTTCTCTGA 202
Db 203 ACTGTGCTGTATATATACCAAAAGAGGATGAGTCGGAGATGGTGT---TGAGGCCAAGA 259
QY 203 AGGAGAACCCCTCGGTGAAGGTGAGGACATCTTCGTCTGCACCAAGGTGTGGAACACC 262
Db 260 TCAAGGAGGGCGTTGTCAAGCGTGAGGATCTGTTTCATCACCAGCAAACTGTGGAACACTT 319
QY 263 TCCACCGTTATGAGGACGCTCTGCTCCATTCAGGACTCCCTGAGCGTCTTGGACTTG 322
Db 320 TCCATCGCCCGGATCTTGTCAAGTCGGCATTTGGAGAACATTTGAGTCCCTTGAGCTGA 379
QY 323 ACTAGCTTGATATGTTCTCTGTTCTACCTGCCCCCATTTCTGCCGAGAGAAATGCCAGGTTG 382
Db 380 AGTACTCGATCTGTACTTATCCACTGCCCCATGGGTCTACAAGGAG-----G 427
QY 383 AGCCCAAGATTTGGCCCTGACGSCAATACGTCTTCAAGGACTGACCGAGACCCCG 442
Db 428 GATGGATCTGTGTCCCAAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 487
QY 443 AGCCCAAGATTTGGCCCTGATGAGGAGAGATTTATGAGGATCGCAAGGCTGAGTCCATTTGTG 502
Db 488 TCGACACGTGNAAGGCCATGAGAGAGTTGGTGGAGAGGGTCTGGTCAAGTCCATTTGTG 547
QY 503 TCTCCAACTGGACCATTTGCCGACCTTGAGAGATGTTCCAAAGTTCGCCAAGGTGATGCTTC 562
Db 548 TTTCCAACTTCAACAGAGGAGATCGAGCGCGTCTTGAGGTGGCCACTATTCCACCAG 607
QY 563 AGCCCAAGATTTGCCGACCTTGAGAGATGTTCCAAAGTTCGCCAAGGTGATGCTTC 622
Db 608 TAACCAATCAGATTGAGTGGCCATTCATATCTGACCAAGAGAGCTGATGACTTCTGCA 667
QY 623 TCTCCAAAGACATTAATGCCCCGCTTCTCTCTCTGGGCTCGCAGAAC 673
Db 668 AGTCAAGGACATTAATCAATCACTGCTACAGTCCCTTGGGATCTCCCAACC 718

RESULT 14
CA015279
LOCUS
DEFINITION
HT13N05r HT Hordeum vulgare subsp. vulgare cDNA clone HT13N05
5-PRIME, mRNA sequence.
CA015279
CA015279.1 GI:24292623
EST.
Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 638)
Zhang, H., Potokina, E., Michalek, W., Weschke, W., Stein, N. and
Graner, A.
Barley ESTs from germinating seeds
```

389 ACCCTGGCCGAGCTTGGCTCGACTACCTTGACCTTACCTCATCTCCACTGCGCCGTTGCC 448
QY
361 GCCGAGAGATGCGCAGGTTGAGCCCAAGATTGGCCCTGACCGCAATACGTCATTCTC 420
Db
449 TTTGCTCCCGCGCGAGCTTCTCCCAAG---TCCGAGGACGCTCCGAGGTGCGACTC 505
QY
421 AAGACCTGACCGAGAACCCCGAGCCCAATGCGCGCTATGGAGAGATTATGAGGAT 480
Db
506 AA--CCCAAGATGTCTCATTTGTCAGACCTGGAAGCCCATGACCGAGCTGCCCAAGTC- 562
QY
481 CGCAGGCGAGTGCATGTTGCTCTCCAACTGAGACCATTCGCGACCTTGAGAGATGTC 540
Db
563 --CAGGTCCCTCCGTCGGTGTCTCCAACTTTACCATTTAGACCCCGACGCGCGCATC 620
QY
541 AAGTTGCGCAAGGTGATGCTCAGCCCAACAGATCGAGATTTCACCCCTTCCTGCCCAAC 600
Db
621 GAGGCCACCGCGTCTGTCGCGCGCTAAACAGATCGAGCGCACACCCCGCTCCCAAC 680
QY
601 GAGGAGTGTGCGAGTACTGTTCTCCAGAACATTAT 638
Db
681 CAGCCCTGTCGACTACTGCGCAAGAACGGCATCAT 718

RESULT 11
LOCUS CB899137 823 bp mRNA linear EST 02-JUL-2003
DEFINITION trico17xh08 T.reesei mycelial culture, Version 3 april Hypocrea
jecorina cDNA clone trico17xh08, mRNA sequence.
ACCESSION CB899137
VERSION CB899137.1 GI:30113795
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
REFERENCE
AUTHORS Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
Olivares, H.A., Teunissen, P.J., Yao, J., and Ward, M.
TITLE Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
JOURNAL J. Biol. Chem. 278 (34), 31988-31997 (2003)
MEDLINE 22803314
PUBMED 12788920
COMMENT Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: Pforeman@genencor.com
Seq primer: LT-F1 primer.
Location/Qualifiers
1..823 /organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trico17xh08"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 3 april"
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Query Match 16.4%; Score 160.4; DB 14; Length 823;
Best Local Similarity 57.7%; Pred. No. 1.1e-24;
Matches 368; Conservative 0; Mismatches 256; Indels 14; Gaps 4;
QY 1 ATGTCTACGGAAGACTTTACATTGAGCAACGCGTCAAGATTCTCGGTCGGTTT 60
Db 149 ATGTCTTCGGAAGACCGTCAACCTTCAACCGGCTACAGATCCCAAGATCGGTAC 208

QY 61 GTTACCTTCCTAGTGAAGTTTCCAAAGCGGAGAACTATATCTGCTGTCACTACTGCCCTG 120
Db 209 GGACCT-----GGAGCGCCTCCGGCGAGGTGCGGGCTGTGTCTTTTGGAGCCCTC 262
QY 121 AAGACGGTTACCGTCACTTGGACTGTGCTGGTACTACCTGAACGAGGGTGAAGTTGGT 180
Db 263 AAGTTGGTACTACCGGCACCTCGACCTGGCCAAAGTCTACGGCAACAGAAAGAGTTGGT 322
QY 181 GAGGTATCCGTGACTTCTCTGAAGGAGAACCCCTCGGTGAAGCGTGAGGACATCTTCGTC 240
Db 323 GAGGGATCAAGAGGCTCTTCTGAGGTCCCGCCCTGAAGCGGAGGATATTTTATC 382
QY 241 TCACCAAGTGTGAACCACTCCACCGTTATGAGGAGCTCTCTGTTCCATTGACGAC 300
Db 383 ACTCTCAAGCTGTGAACAACCTCCCAAGCCGAGGACGTCGAGCCGCTCTCGACGAC 442
QY 301 TCCTCTGAAGCGTCTGGACTTGAAGTGTGATATGTTCTCTGTTCACTGGCCCAATTGCT 360
Db 443 ACCCTGGCGAGCTTGGCCTCGACTACCTTTGACCTCTACCTCATCGACTGGCCCTTGGC 502
QY 361 GCCGAGAGAATGGCCAGGTTGAGCCCAAGATTGGCCCTGACGGCAAAATACGTCATTCTC 420
Db 503 TTTGCTCCCGCGCGACCTCTTCCCAAG---TCGAGGACGGCTCCGAGGTGCGAGTCTC 559
QY 421 AAGGACCTGACGAGAACCCCGAGCCACATGCGCGCTATGGAGAAGATTATGAGAT 480
Db 560 AA--CCCAAGATGTGTCCATTGTCCAGACCTCGAAGGCCATGACGAGTGGCCCAAGTC- 616
QY 481 CGCAAGCGCCAGTTCATTGTTGTTCTTCAACTGGACCATTTGCCGACTTGAGAAGATGTC 540
Db 617 --CAGGTCCGTCGCTGCTCTTCCAACTTTACCATTTAGCAACCCCGCCGATC 674
QY 541 AGTTGCGCAAGTGTATGCTTCAACCAACAGATCGAGATTCAGCCCTTCCTGCCCAAC 600
Db 675 GAGGCCACCGCGCTGCTGCTCCCGCCGTTAAACACAGATCGAGCGCACACCCCGCTCCCAAC 734
QY 601 GAGGAGCTGTGCGAGTACTGTTCTTCCAAAGACATTAT 638
Db 735 CAGCCCTCATCGACTACTTGGCCCAAGAACGGCATCAT 772

RESULT 12
LOCUS AA785751
DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
Location/Qualifiers

AA785751 189 bp mRNA linear EST 31-JUL-1998
hlh04a1.x1 Aspergillus nidulans 24hr asexual developmental and
vegetative cDNA lambda zap library Emericella nidulans cDNA clone
hlh04a1 5', mRNA sequence.
AA785751
AA785751.1 GI:2845919
EST.
Emericella nidulans (anamorph: Aspergillus nidulans)
Emericella nidulans
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Emericella.
1 (bases 1 to 189)
Kupfer, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R.,
Prade, R., and Roe, B.
An Aspergillus nidulans EST Database
Unpublished (1998)
Other_ESTRs: hlh04a1.f1
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: SK
High quality sequence stop: 182.
Location/Qualifiers


```
Db 265 ACCGAAAGCTTTGGAACACAAACACCGCCCTGAGGAGGTGCTGTGCTCTTGAGGAC 324
Qy 301 TCCTGTGAAGCGCTTGACTGACTAGCTGATGTTCTCTGTTCACTGGCCCATGCT 360
Db 325 ACCCTCAGGAGCGCTGCTGTTGAGTATCATGACCTCTGTTGATCCACTGCGCTGTGCT 384
Qy 361 GCCAGAGAATGCCAGGCTGAGCCCAAGATTGGCCCTGACGCGCAATACGTCATCTC 420
Db 385 TTCAAGAACGAAACACAGCTTTTCCCACTGAAGAGGGCGATGACGGCAAGACTGCTCTT 444
Qy 421 AAGCACTGACCGAGAACCGGAGCCACATGCGCGCTATGAGGAAGATTATGAGGAT 480
Db 445 GACCAGGAGGTGACCCCTCTCCAGACCTGTGAGGCGCTGACCAAGCTGCCCAAGGAG-- 501
Qy 481 CGCAAGCGCAGGTCATGTTGTTCTCAACTGGACCAATTGCCAGCTTGAGAGAGTGTC 540
Db 502 ---AAGTTGCTCTCAATGTTGTTCTTCACTTCAAGAGAGATGTTGAGCAATCATC 558
Qy 541 AAGTTCGCCAAGGTCATGCTCAGCCCAACAGATGAGATCACCCCTTCGTCGCCAAC 600
Db 559 AAGGACACTGCGCTCAGCCCGCCATGAACAGATCGAGCGCAACCCCGCTTCTCTCA 618
Qy 601 GAGGAGCTGTGCTAGTACTGTTCTTCCAAAGCAATATGCGCGTGGCTACTCTCTCTG 660
Db 619 CTTGAGCTTGTCAGTACTTCCAGGAGAGGTTACTTCTCTACTGCTTACTCGCCTTC 678
Qy 661 GGCTCGCAGAACACAGG 676
Db 679 GGCAACACAGCTGGG 694

RESULT 8
CNS06060 1087 bp DNA linear GSS 05-JUN-2001
LOCUS T7 end of clone AW0AA003D07 of library AW0AA from strain CLIB 89 of
DEFINITION Yarrowia lipolytica, genomic survey sequence.
ACCESSION AL410278
VERSION AL410278.1 GI:12178444
KEYWORDS GSS.
SOURCE Yarrowia lipolytica
ORGANISM Yarrowia lipolytica
REFERENCE 1 (bases 1 to 1087)
AUTHORS Souciot,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Maupertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEMS Lett. 487 (1), 3-12 (2000)
20584711
11152876
2 (bases 1 to 1087)
AUTHORS Casaregola,S., Neuveglise,C., Lepingle,A., Bon,E., Feynerol,C.,
Artiguenave,F., Wincker,P. and Gaillardin,C.
Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia
lipolytica
FEMS Lett. 487 (1), 95-100 (2000)
20584727
11152892
3 (bases 1 to 1087)
AUTHORS Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
```

```
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared, and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
Location/Qualifiers
1..1087
/organism="Yarrowia lipolytica"
/mol_type="genomic DNA"
/strain="CLIB 89"
/db_xref="taxon:4952"
/clone="AW0AA003D07"
/cloned_lib="AW0AA"
/note="end : 17"
<23..>763
misc_feature
/note="similar to Saccharomyces cerevisiae ORF YDR368w [
IPRI ; strong similarity to members of the aldo/keto
reductase family]"
/evidence=not_experimental
<23..>733
misc_feature
/note="similar to Saccharomyces cerevisiae ORF YOR120w [
GCY1 ; galactose-induced protein of aldo/keto reductase
family]"
/evidence=not_experimental
ORIGIN
Query Match 17.0%; Score 166; DB 29; Length 1087;
Best Local Similarity 56.6%; Pred. No. 7, 7e-26;
Matches 379; Conservative 0; Mismatches 270; Indels 21; Gaps 3;
Qy 223 CGTGAGGACATCTTCTGTCGACCAAGGTGTGGAACCACTCCACCGTTATGAGACCTC 282
Db 83 CGAGAGGAGATCTTCTGACCAACAAACTGGAACAACGCCCGAACCAAGAGAGGCGCTC 142
Qy 283 CTCGTGCTCATTGACGACTCCCTGTAAGCGTCTTGACTTGACTGACTGATGATGCTCCTC 342
Db 143 AAGAAGCAGCTCGACACAGTCCCTCAAGGAGCTGCAGACCGACTAGCTGATGCTGCTC 202
Qy 343 GTTCACTGCGCCCATTTGTCGCGAGAGAATGCCAGGTGAGCCCAAGATTGGCCCTGAC 402
Db 203 ATCCACTGCGCCGTTGTTTTC---AAGACCGCGGAGGAGCTCTTCCCCACAAACCCGAC 259
Qy 403 GGCATATGCTCATCTCAAGGACCTGACCGAGACCCGAGCCACATGGCGGCTATG 462
Db 260 GGCATGTCGCCCTTGGCGAGTCCCTGTGAGG-----GACTTGGAGGCTTTC 310
Qy 463 GAGAAGATTATGAGGATCGCAAGCCAGGTCCATTGTTGTTCTCAACTGGACCATGCTC 522
Db 311 GAGGAGCTCGTGAAGAGGAGGCAAGGTCAAGTCCATCGGTCTTCCAACTTCAACCGAGGAG 370
Qy 523 GACCTTGAGAAGATGTCGAAGTTCGCAAGTTCATGCTCAGCCACCAACAGATCGAGATT 582
Db 371 CGAATTGAGAAGCTCTCTCAAGACCCCAAGATTCCTCCCTGCGCTCAACAGATTGATAC 430
Qy 583 CACCCCTTCTCTGCCCAAGCAGGAGCTGTGTCAGTACTGCTTCTTCCAAAGACATTATGCC 642
Db 431 CACCCCTACTCTGACGACGCGGCTCTGCGCAAGTACCTTGGAGCAAGAACATTGCTCTC 490
Qy 643 GTGGCTACTCTCTCTGCGCTCGCAGAACACAGGTTCCACACCGGTCAGCGGTGAGGGGTGAC 702
Db 491 GAGGCTTACTCTCTCTCGTTAAACAACACTTACGGCTTTCCTCC-----CGAGCTATT 541
Qy 703 GAGAACAGACTCTGAACAGAGATCGCCGAGAGGGCGCAACACCTTGTCTCAGGTCTT 762
Db 542 GATGACGAGGAGTATCGATCTGCGCAAGAGAGGGGAGGATGCTGCTGCCCTCAT 601
Qy 763 ATGCTGCTGGGTCTGCGCGCTGGCTGCTGCTGCTTCTCCCAAGAGTCCCAACCCCAAGCGC 822
Db 602 GTCAACTGGATCAGATCCAAAGAACCATGTTGTTCTGCTCAAGTCCGTCACCCCTTCTCGA 661
Qy 823 ATTGAGTCCAACTTCAAGAGCATTTGAGCTTCTCCGTCGCGACTTTTGAAGCCATCAATGCC 882
```

Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu

Contact Dr. Marian Beremand regarding clone availability Included
is the best homolog from a blastx search of Genbank nr 04-09-01
526 2e-53 gi|3724334|dbj|BAA33 (AB014493) reductase [Gibberella
zeae]

Seq primer: T3
High quality sequence stop: 130.

FEATURES

Location/Qualifiers
1..390
/organism="Fusarium sporotrichioides"
/mol_type="mRNA"
/strain="Tri 10"
/db_xref="taxon:5514"
/clone="nd1015"
/clone_lib="Fusarium sporotrichioides Tri 10 overexpressed
cDNA library"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; 5' end of cDNA cloned into EcoRI site of
pBluescript; 3' end of cDNA cloned into XhoI site of
pBluescript"

ORIGIN

Query Match 19.3%; Score 188.6; DB 12; Length 390;
Best Local Similarity 73.1%; Pred. No. 4.7e-31;
Matches 242; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 2 TGTCTACGGAAGACTTTCACATTGAGCAACGGCTCAAGATTCCTGGGTGGCTTTC 61
Db 56 TTTCTAACAGACTGTCTTACCTCAACACGGTCAAGATTCCTGGGTGGCTTTC 115
QY 62 GTACCTTCGCTGAGAGTTTCCAAAGCGCGAGACCTATCTGTGTCCACACTGCCCTGA 121
Db 116 GCACCTTGTCTAAGAGGGCGCCAAAGCGCGAGACTTACAGGCTGTCTATTCGCCCTCA 175
QY 122 AGACCGGTTACCGTCACTTGACTGTGCTGGTACTACCTGAACGAGGTTGAGTTGGTG 181
Db 176 AGACCGGTTACCGTCACTTGACTGTGCTGGTACTACCTGAACGAGGTTGAGTTGGTG 235
QY 182 AGGTTATCGTGACTTCTCTGAAGAGAACCCCTCGGTGAAGCGGTGAGGACATCTTGTCT 241
Db 236 AGCGGTGCGTGAATTTCTGCCGAGAGCAACGGCTCAAGCGGAGGATCTTTATCT 295
QY 242 GCACCAAGGTTGGAACCACTCCACCGTTATGAGGACGTCCTCTGTCTCACTGACGACT 301
Db 296 GCACCAAGGTTGGAACCACTCCACGAGCTGAGGAGGTCAAGTGGTCACTTCGAGAACT 355
QY 302 CCTGTAGCGTCTTGGACTTGACTACCTTGA 332
Db 356 CTCCTCAAGACTTTGGCGCTGACTACATTGA 386

RESULT 7
BI948347
LOCUS BI948347 1061 bp mRNA linear EST 19-OCT-2001
DEFINITION HVSM10009C04f Hordeum vulgare spike EST library HVCNDA0012
(Fusarium infected) Hordeum vulgare subsp. vulgare cDNA clone
HVSM10009C04f, mRNA sequence.
ACCESSION BI948347
VERSION BI948347.1 GI:16288644
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 1061)
Wing, R., Muehlbauer, G.J., Close, T.J., Kleinhofs, A., Wise, R.,
Heinen, S., Begum, D., Yu, Y., Henry, D., Palmer, M.,
Rambo, I., Simmons, J., Fenton, R.D., Malatrasi, M., Choi, B.W.,
Oates, R. and Main, D.
Development of a genetically and physically anchored EST resource

JOURNAL COMMENT

for barley genomics: Fusarium infected Morex spike cDNA library
Unpublished (2001)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 541
Seq primer: AATTAACTCTCACTAAAGGG
High quality sequence stop: 785.

FEATURES

Location/Qualifiers
1..1061
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSM10009C04f"
/tissue_type="Spike"
/lab_host="TJCL1"
/clone_lib="Hordeum vulgare spike EST library HVCNDA0012
(Fusarium infected)"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; Plants were grown at the University of Minnesota in
the GJ Muehlbauer lab; spikes were harvested and snap
frozen at 0, 1, 2, 3, 4, 5, 6, and 8 days after Fusarium
graminearum inoculation (Heinen). In the TJ Close lab at
the University of California, Riverside, total RNA was
prepared from each sample pool, equal quantities of all
eight RNA pools were combined, poly(A) RNA was purified
from the mixture, one primary unamplified cDNA library was
made, and 1 million pfu were in vivo excised to give
pBluescript SK(-) cDNA phagemids (Choi, Fenton,
Malatrasi). Phagemids were plated and picked at the
Clemson University Genomics Institute (CUGI) (Begum,
Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,
Oates, Rambo, Main). The sequence has been trimmed to
remove vector sequence and contains a minimum of 100 bases
of phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html)"

ORIGIN

Query Match 18.0%; Score 176; DB 12; Length 1061;
Best Local Similarity 56.1%; Pred. No. 4.8e-28;
Matches 379; Conservative 0; Mismatches 285; Indels 12; Gaps 2;
QY 1 ATGCTTACGGAAGACTTTCACATTGAGCAACGGCTCAAGATTCCTGGGTGGCTTTC 60
Db 31 ATGCTTTCGGTCGAACTGTCTACTCTCACTCGGCTCGAAGATCCCCCAGATCGGCTAC 90
QY 61 GGTACTTTCGTTAGTGAAGTTTCCAAAGCGCGAGACTATCTGTGTCCACACTGCCCTG 120
Db 91 GGACAT-----GGCAAGCTCTCCCGTGGGTGGCAATGTGTCTACGAGGCCCTT 144
QY 121 AAGACCGGTTACCGTCACTTGACTGTGCTGGTGTACTACTCTGAACGAGGTTGAGTTGGT 180
Db 145 AAGGCTGTGCTACCGTCACTTGTCTCCCAAGATCTTACCAAGAACCCAGCGTGGTTGGT 204
QY 181 GAGGTTATCCGTGACTTCTCTGAAGAGAAACCCCTCGGTGAAGCGGTGAGGACATCTTCGTC 240
Db 205 GAGGTTATCAAGAGGCCCTCAGCGAGTCCCTGGCTTAAGCGGTGAGGACATCTTCATC 264
QY 241 TGCACCAAGGTTGGAACACCTCCACCGTTATGAGGAGCTCTCTGTGTCATTGACGAC 300


```

QY      313  CTTGGACTTGACTAGTTGATATGTTCTCTGTTCACTGGCCCAATGCTGCCGAGAAGAAT 372
DB      463  CTAAGGTCGATTACATTGACCTGTTCTCTGTTCACTGGCCCAATGCTGCCGAGAAGAAT 522
QY      373  GGCAGAGGTGAGCCCRAGATTGGCCCTGAGCGGCAATACATCTCTCAAGGACCTTGACC 432
DB      523  AGCGACAGGAGGTCAAGCTGGGGCCCCGATGGCAAGTATGTCATCAACCAAGCCCTCGACG 582
QY      433  GAGAACCCCGAGCCACATCGCGGCTATGGAGAAGATTTATGAGGATCGCAAGGCCAGG 492
DB      583  GAAACCCAGAGCCCAACATGCGAGCCATGGAAGAGCTTGTGAAGCGSCCTCGTCAAG 642
QY      493  TCCATTGGTGTCTCAACTGACCAATGCGGACCTTGAGAAGATGTTCCAAAGTTGCGCAAG 552
DB      643  GCAATTGGAGTATCAACTGCGACGATNCGGGGTGAAGAAGCTCTTCAGATCGCCCAAG 702
QY      553  GTCATGCTCAAGCCCAACCAAGATCGAGATTCA 584
DB      703  ATCAAGCGCGCAGTGAACCATGATGAAGATCA 734

RESULT 3
CB897574
LOCUS   trico011xm03 T.reesei mycelial culture, Version 3 april Hypocrea
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Hypocrea jecorina (anamorph: Trichoderma reesei)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE
1 (bases 1 to 792)
FOREMAN,P.K., BROWN,D.E., DANKMEYER,L., DEAN,R., DIENER,S.,
DUNN-COLEMAN,N.S., GOEDEGEBOUR,F., HOUFEK,T.D., ENGLAND,G.J.,
KELLEY,A.S., MEERMAN,H.J.J., MITCHELL,T., MITCHINSON,C.,
OLIVARES,H.A., TEUNISSEN,P.J., YAO,J. and WARD,M.
Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
J. Biol. Chem. 278 (34), 31988-31997 (2003)
22803314
MEDLINE
PUBMED
12788920
COMMENT
Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: Pforeman@genencor.com
Seq primer: LT-F1 primer.
Location/Qualifiers
1..792
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/db_xref="taxon:51453"
/clone="trico011xm03"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 3 april"
/note="vector: pREP3; Site 1: Not 1/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

FEATURES
source
Query Match 27.1%; Score 265; DB 14; Length 792;
Best Local Similarity 66.3%; Pred. No. 9.9e-48;
Matches 379; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY      13  AAGACTTCACATTGAGCAACGGCGTCAAGATCTCTGGCTCGGCTTGTGACTTCGCT 72
DB      221  AAGAGCTACACTCTGAACACCGGTGCCAAGATTCCTCGGTCGGGTTCGGCACATTCGCC 280

```

```

QY      73  AGTGAGGTTCCAAAGGGCGAGACCTATATCTGCTCCACCTGCGCCCTGAAGACCGGTTAC 132
DB      281  AATGAGGTTGCCAAGGGCGAGACATACGACATGTTTACAAAGGCACCTGGAGTTGGATAC 340
QY      133  CGTCACTTGGACTGTGCTCTGTTACTTACCTGAACAGAGGTGAGGTTGGTGGAGGTTATCCGT 192
DB      341  CGCCACCTTGATTGGCGCTGGTTTACCAACAGAGATGAGGTTGGTGGACGCGTACGC 400
QY      193  GACTTCTTGAAGGAGAACCCCTCGGTGAAGCGTGAAGACATCTTCGTCTGCACCAAGGTG 252
DB      401  GATTTCTCGCCGCGGACCGGACGCTGAACCGCGAGGATCTCTTCATTGACCCAAAGTT 460
QY      253  TGGAAACCACTCCACCGTTATGAGGACCTCTCTGTTCCATTGACGACTCCCTGGAAGCGT 312
DB      461  TGGAAACCACTTCGTCATGAGCCAGAGACGTCAGTGGAGCGCCCAAGAACTCGTGGCAAAAC 520
QY      313  CTTGAGACTTGACTACGTTGATATGTTCTCTGTTCACTGGCCCAATGCTGCCGAGAAGAAT 372
DB      521  CTCGAAGGTCGATTACATTGACCTGTTCTCTGTTCCACTGGCCAATCGCGCGCGAGAAGAAC 580
QY      373  GGCAGAGGTGAGCCCAAGATTGGCCCTGACGCGCAAAATACGTCATTCTCAAGCACTGACC 432
DB      581  AGCGACAGAGCGTCAAGCTGGCCCCGATGCGAAGTATGTCATCAACCAAGCCCTGACG 640
QY      433  GAGAAACCCCGAGCCCACTGCGGCTATGAGAAGATTTATGAGGATCGCAAGGCCAGG 492
DB      641  GAAACCCAGAGCCCAACATGCGAGCCATGGAAGAGCTTGTGAAGCGGCTCTGTTCAAAG 700
QY      493  TCCATTGGTGTCTCAACTGAGACCAATGCCGACCTTGAGAGATGTCACAGTTTGGCCCAAG 552
DB      701  GCAATTGGAGTATCAACTGAGACGATNCGGGGTGAAGAAGCTCTTCAGATCGCCCAAG 760
QY      553  GTCATGCTCAAGCCCAACCAAGATCGAGATTCA 584
DB      761  ATCAAGCGCGCAGTGAACCATGATGAAGATCA 792

CF866105      704 bp mRNA linear EST 31-OCT-2003
trico05xa21.bi T.reesei mycelial culture, Version 6 October 2003
Hypocrea jecorina cDNA clone trico05xa21, mRNA sequence.
CF866105
CF866105.1 GI:39120731
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Hypocrea jecorina (anamorph: Trichoderma reesei)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE
1 (bases 1 to 704)
DIENER,S.E., DANKMEYER,L., DUNN-COLEMAN,N., HOUFEK,T.D.,
MITCHELL,T.K., VAN SOLINGEN,P., TEUNISSEN,P.J.M., WARD,M. and
DEAN,R.A.
Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset
Unpublished (2003)
CONTACT: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: LT-F1 primer.
Location/Qualifiers
1..704
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/db_xref="taxon:51453"
/clone="trico05xa21"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October

```

Email: r.bruskiewich@giar.org
 International Rice Information System (IRIS;
<http://www.iris.irri.org>): D0200799
 Assignment of putative function to the sequence by S. Rudd of the
 Munich Information Center for Protein Sequences
 (<http://mips.gsf.de>)
 Plate: 03 row: B column: 08.
 Location/Qualifiers
 1..886
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="IR64"
 /db_xref="taxon:39946"
 /clone="C0000800"
 /tissue_type="panicles"
 /dev_stage="Flowering"
 /clone_lib="IRRI Drought Stress Panicle Library"
 /note="Vector: pbluescript II SK+; Water stress was
 applied by not watering for 4 consecutive days. Panicles
 were collected from control (well watered) and stressed
 plants at 2 days before heading, at heading, 50% flowering
 and 4 days after 50% flowering."

FEATURES

source

Query Match 36.1%; Score 353.2; DB 14; Length 886;
 Best Local Similarity 68.2%; Pred. No. 3.8e-67;
 Matches 527; Conservative 0; Mismatches 242; Indels 4; Gaps 3;

QY 5 CTACCGAAGACTTTCACATGTAGCAACGCGCTCAAGATTCCTGGCGTTGGTA 64
 DB 114 CAACACGAGACATACAAAGCTCAACATGCGTTGAGATTCCTCCGCTAGGATTCGAA 173
 QY 65 CTTTCGCTAGTGAAGTTCCAGGCGGAGACCTATCTGCTGCACACTGCCCTGAAGA 124
 DB 174 CATTTCGCAACGAGGCTCCAGGCGGAGACCTACGCTGCCGTACACATGCTCTGAAGA 233
 QY 125 CCGTTACCGTCACTTGGACTGTCCCTGGTACTACCTGAACGAGGTTGAGTTGGTAGG 184
 DB 234 CTGGCTACCGACACTCTGACTGCGATGTTCTACCAAGACGAGGCGAGTGGTGGAGG 293
 QY 185 GTATCCGTGACTTCTGAGAGAGACCTCGTGGTGAAGCGTGAAGGATCTTCTGTGCA 244
 DB 294 CCGTCCGCGACTTCTTCGCGCCAGAACTCCCTCCGCTCATCTCGCAAGGATCTTCACTGCA 353
 QY 245 CCAAGGTGTGAACACACCTCCACCGCTTATGAGGACGCTCTCTGCTCCATTGAAGACTCCC 304
 DB 354 CAAGGTCTGAACCACTCTCCAGCGCCGAGGAGTAGATGTCTGTTCAACAGCTCGC 413
 QY 305 TGAAGCGTCTTGGACTTGACTAGCTTGATGTTCTCTGTTCACTGGCCCATTTGCTGCCG 364
 DB 414 TCGAAGTTTCAGATGGACTACATCGATCTCTTTCTCGTCACTGGCCCTATTGCTGCTG 473
 QY 365 AGAAGATGGCCAGGTTGAGCCCAAGATTGGCCCTGACGCAATACGTATCTCAAGG 424
 DB 474 AGAAGACGAGATACATATGCTCAAGCTTGAGCTGATGCACTATCATCAAGAGG 533
 QY 425 ACCTGACCGAAGACCCGAGCCCACTATGCGCGCTATGAGAGAGATTATGAGGATCGCA 484
 DB 534 AGCTGACCGAAGACCTGAGCCCACTTGGCGCGCAATGGAGGCCATCTACAACCTCTGGCA 593
 QY 485 AGGCGAGTCCATTGGTGTCTCCAACTGGACCATTTGCCGACTTGAGAAGATGTCGAAGT 544
 DB 594 GGGCTAAGGCATATGGTGTGTCCTCACTGGACCGCTCAAGGACNTCAAGAGTCTAGCT 653
 QY 545 TCGCCAAAGTTCATGCTCAC -GCCAACGAGATCGAGATTACCCCTTCTCGCCCAAGCAG 603
 DB 654 TCGCAAGGTCAAGNCCGCGAGTCAACCAATCGAGATCCGATCCCTTCTCTTAAGGCC 713
 QY 604 GAGTGTGTGAGTACTGCTTCTCAAGACGATTTATGCCGTGGCTACTCTCTCTGGGC 663
 DB 714 GAGCTCGNCAAGTACTGCGACGAGACGAGATTTCTTCAGCGGCTACTTGGCGGTTGGG 773
 QY 664 TCGCAGAACCAAGTTTCCCAACCGGTTGAGCGGCTCAGCGAGAACAGACTCTCTGAACGAG 723

DB 774 TTCCANANACAGTCCCA-CACAGCGAGAGGTCGCAACCAACCCAGCTC--AACGAG 830
 QY 724 ATCGCGGAGAGGGCGGAACACCTTCTCAGGTTCTTATTGCTCTGGGGTCT 776
 DB 831 GTTCTGAGCGGACGCGCACGATCTCGACAGTCTGCTGATGGGCTTT 883

RESULT 2
 CF867536 734 bp mRNA linear EST 31-OCT-2003
 LOCUS tricol1xm03.b1 T.reesei mycelial culture, Version 6 October 2003
 DEFINITION Hypocrea jecorina cDNA clone tricol1xm03, mRNA sequence.
 ACCESSION CF867536
 VERSION 1
 KEYWORDS EST.
 SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
 ORGANISM Hypocrea jecorina
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocryomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
 REFERENCE 1 (bases 1 to 734)
 AUTHORS Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,
 Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and
 Dean,R.A.
 TITLE Analysis of the protein processing and secretion pathways in a
 Trichoderma reesei EST dataset
 JOURNAL Unpublished (2003)
 COMMENT Contact: Ralph A. Dean
 Fungal Genomics Laboratory
 North Carolina State University
 Campus Box 7251, Raleigh, NC 27695, USA
 Tel: 919-513-0020
 Fax: 919-513-0024
 Email: ralph_dean@ncsu.edu
 Seq primer: LT-F1 primer.
 Location/Qualifiers
 1..734
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QM6a"
 /db_xref="taxon:51453"
 /clone="tricol1xm03"
 /dev_stage="mycelia"
 /clone_lib="T.reesei mycelial culture, Version 6 October 2003"
 /note="Vector: pRBP3v; Site 1: Not I/Sal I; Mycelial
 culture grown from 24 hrs to 6 days with varying Carbon
 and Nitrogen sources and concentrations."

ORIGIN

Query Match 27.1%; Score 265; DB 14; Length 734;
 Best Local Similarity 66.3%; Pred. No. 9.5e-48;
 Matches 379; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 13 AGACTTTCACATTGAGCAACGCGCTCAAGATTCCTGGCTGGCTTTGGTACCTTCGCT 72
 DB 163 AAGAGCTACACTCTGAACACCGGTGCCAAGATTCCTCCGGTTCGGGTCATTCGCC 222
 QY 73 AGTGAAGGTTCCAAAGGGGAGACCTATCTACTGCTGCACACTGCCCTCAAGACCGGTTAC 132
 DB 223 AATGAGGTTGCCAAGGGCGAGACATACGAGCTGTTACAAAGGCACTGACGTTGATAC 282
 QY 133 CGTCACTTGGACTGTGCTGTGTTACTACCTGAACGAGGTGAGTTGGTGAAGGTATCCGT 192
 DB 283 CGCCACCTTGTATGTCGCTGTTTACCACACGAGATGAGTTGGTGAACGCGGTACGC 342
 QY 193 GACTTCTCAAGGAGAACCCCTCGGTGAAGCGTGAAGACATCTTCTGTGACACCAAGGTG 252
 DB 343 GATTTCTCGCCCGCCGCCACCGCTGAACCGAGGATCTTCTGATTTGACCAAGTT 402
 QY 253 TGAACACCACTCCACCGTTTATGAGACGTCCTCTCTGTTCAATTGACGACTCCCTCTGAAGCT 312
 DB 403 TGAACACCACTGTCATGAGCGAGGACGTCAGTGGAGCGCAAGAACTCTGTCGCAAAAC 462

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2004, 11:17:03 ; Search time 2523 Seconds
(without alignments)
11575.590 Million cell updates/sec

Title: US-10-004-115B-2

Perfect score: 978
Sequence: 1 atgtctaacggaagacttt.....ccagaacctgtctggtga 978

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estopl:*

7: em_estro:*

8: em_estc:*

9: gb_est1:*

10: gb_est2:*

11: gb_estc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	353.2	36.1	886	14 CA764322	CA764322 AF53-Rpf
2	265	27.1	734	14 CF867536	CF867536 tricol1xm
3	265	27.1	792	14 CB897574	CB897574 tricol1xm
4	220.8	22.6	704	14 CF866105	CF866105 tricol005xa

5	220.8	22.6	757	14	CB896064
6	188.6	19.3	390	12	BI200418
7	176	18.0	1061	12	BI948347
8	166	17.0	1087	29	CNS06060
9	164.8	16.9	534	9	AA786853
10	160.4	16.4	769	14	CF869025
11	160.4	16.4	823	14	CB899137
12	152.2	15.6	189	9	AA785751
13	149	15.2	779	9	AA263299
14	148.8	15.2	638	13	CA015279
15	148.6	15.2	758	9	AA697426
16	139.4	14.3	856	12	BI948257
17	138.6	14.2	449	12	BM872415
18	136.6	14.0	997	11	CNS08080
19	132.2	13.5	827	13	BQ751060
20	131	13.4	1036	11	CNS0808N
21	129.4	13.2	955	11	CNS0945X
22	129.4	13.2	1311	11	AY105383
23	128	13.1	643	13	BU645057
24	128	13.1	821	29	CC594922
25	127.8	13.1	1021	11	CNS0945X
26	127	13.0	790	14	CF642818
27	124.4	12.7	706	14	CF644498
28	122.2	12.5	770	13	EX607579
29	120.2	12.3	1201	13	EX417080
30	119.8	12.2	668	12	BI619770
31	119.6	12.2	1201	13	EX417343
32	119.4	12.2	644	10	BE974991
33	119.4	12.2	665	12	BI623282
34	119.4	12.2	666	12	BI485421
35	119.2	12.2	1019	13	EX331589
36	119.2	12.2	1118	13	EX397710
37	119.2	12.2	1179	13	EX377469
38	118.4	12.1	639	9	AI114144
39	118	12.1	563	12	BI369569
40	117.8	12.0	1044	12	BM561410
41	117.4	12.0	610	12	BG368501
42	117.2	12.0	1017	29	CNS0784Q
43	116.8	11.9	809	12	BI857430
44	116.6	11.9	1176	9	AL547253
45	116.4	11.9	640	9	AI135240

ALIGNMENTS

RESULT 1

CA764322

LOCUS

DEFINITION

CA764322

AF53-Rpf 03 B08 T7 018.abi

Oryza sativa (indica cultivar-group)

to Alcohol dehydrogenase [NADP+]

(3-DG-reducing enzyme), mRNA sequence.

CA764322

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

886 bp mRNA linear EST 08-JAN-2003
AF53-Rpf 03 B08 T7 018.abi IRR1 Drought Stress Panicle Library
Oryza sativa (indica cultivar-group) cDNA clone C000800.5, similar
to Alcohol dehydrogenase [NADP+]

(3-DG-reducing enzyme), mRNA sequence.

CA764322

CA764322.2 GI:27546321

EST.

Oryza sativa (indica cultivar-group)

Oryza sativa (indica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoidae; Oryzae; Oryza.

1 (bases 1 to 886)

Bennett, J., Arumugam, K., Lafitte, R., Wen, J., Rudd, S. and

Bruskiewich, R.N.

IRRI Drought Stress Panicle cDNA Library

Unpublished (2002)

On Dec 2, 2002 this sequence version replaced gi:2593577.

Contact: Richard Bruskiewich

Bionetrics and Bioinformatics Unit

International Rice Research Institute

DAPO 7777, Metro Manila, Philippines

Tel: +63-2-845-0563

Fax: +63-2-845-0606

QY 623 TCTCAAGAAACATTATGCCCCGCTGCTACTCTCTCTGGGTGCGAGAA---CCAGGTTTC 679
Db 668 AGTCAAAGGACATTACATCACTGCTACAGTCCCTTGGGATCTCCCAACCGGCCATGGG 727
QY 680 CCACCACCGGTGAGCGGGTCAAG---GAGAACAAAGACTCTGAACGAGATCGCCGAGNAGG 736
Db 728 CCAAGGCTGGGTGATCCCGGTCACTCTAGAGGAGGCTAAGATCAAGGAAATTGCGGCTAAGA 787
QY 737 CGCGCAACACCCCTTGCTCAGGTTCTTATTGCTGGGTCTGCGCGCTGAGTCAAGTCTGTTTC 796
Db 788 AGAAGAGACCCCTGGACAGATCTTATTGATACCAAGGTTCAAGCGTGCCCAACATTGTTA 847
QY 797 TCCCAAGAGCTCAACCCCAAGCGCATTAAGTCCAACTTCAAGAGCATTAAGCTC 852
Db 848 TCCCCAAATCTGTGACCAAGGACCGCATCGAGTCCAACTTCCAGGTTCTCGACTTC 903

RESULT 15

ABZ56044

ID ABZ56044 standard; cDNA; 403 BP.

XX AC ABZ56044;

XX DT 28-MAR-2003 (first entry)

XX DE Aspergillus oryzae polynucleotide SEQ ID NO 5157.

XX KW Aspergillus oryzae; fermentation; fungus; industrial; EST;
XX KW expressed sequence tag; gene; ss.

XX OS Aspergillus oryzae.

XX PN WO200279476-A1.

XX PD 10-OCT-2002.

XX PF 22-MAR-2002; 2002WO-IB000890.

XX PR 30-MAR-2001; 2001JP-00098371.

XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

XX PA (NARE-) NAT RES INST BREWING.

XX PA (NORQ) NAT FOOD RES INST MIN AGRIC.

XX PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;

XX PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;

XX DR WPI; 2003-046817/04.

XX PT Detection of expression of specific Aspergillus genes for monitoring the
XX PT fermentation and growth conditions of the fungus, using DNA probes.

XX PS Claim 1; SEQ ID NO 5157; 48pp + Sequence Listing; Japanese.

XX CC The invention relates to a polynucleotide having any of 6006 specific
XX CC sequences (ABZ5088-ABZ5693), which are expressed by a fungus under
XX CC specific culture conditions including one or more of eutrophic,
XX CC oligotrophic, solid, early germination, alkaline, high temperature, low
XX CC temperature or maltose culture or polynucleotides stringently hybridising
XX CC to these sequences. The polynucleotides are useful for monitoring the
XX CC progress of fermentation and the growth conditions of a fungus,
XX CC especially of Aspergillus oryzae which is widely used in industrial
XX CC fermentation. Also monitoring for fungal contamination. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 403 BP; 103 A; 107 C; 101 G; 92 T; 0 U; 0 Other;

Query Match 18.0%; Score 176.4; DB 7; Length 403;

Best Local Similarity 70.9%; Pred. No. 3e-37;

Matches 248; Conservative 0; Mismatches 101; Indels 1; Gaps 1;

QY 1 ATGTCTAACGGAAAGACTTTTCACATTGAGCAACGGCGTCAAGATTCTCTGGGTCGGCTTTT 60
Db 54 ATGTCTCTCTGGCGTATCCTTCAAAACAAGCAACGGCGTCAACATCCAGGGCGTGGCTTC 113
QY 61 GGTACCTTCGCTAGTGAAGGTTCCAAAGGGCGAGACCTATCTGCTGTACCACTGCCCTG 120
Db 114 GGTACCTTCGCTAAGCAAGGCTCCAAAGGGCGAGACCTAAGGCTGTCCGTCACCGTTG 173
QY 121 AAGACCGGTTACCGTCACTTGGACTGTGCTGGTACTACTGAAACGAGGGTGAGGTTGGT 180
Db 174 AAAGTCGGCTACCGCCACCTCGAATGTGCTGGTTCTATGCCAATGAGGACGAAGTAGGA 233
QY 181 GAGGGTATCCGTGACTTCTCTGAAGGAGAACCCCTCGGTGAAGCGTGAGGACATCTTCGTC 240
Db 234 CAGGGGATTCAGATTTCTCTGAAGAGAACCCCGTCGGTTAAACGGGAGGATCTGTTTGTG 293
QY 241 TGCACCAAGGTTGTGAACCCACCTCCACCGTTATGAGGACGTCCTCTGGTCCATT-GACGA 299
Db 294 ACGACGAAGTGTGAAACCACTGCATCGGTATGAGATGTTCTCTGGTCCGTTGGAAGA 353
QY 300 CTCCTGAAAGCGTCTTGGACTTGAATGAGTGTGATATGTTCTCTCGTTCACT 349
Db 354 TTCAATTGAAGAACTTGCAATTGAGCTACGTGATCTGTTTCTCTCGTCACT 403

Search completed: May 29, 2004, 11:11:56

Job time : 455 secs

AC AAT29160;
 XX 16-OCT-2003 (revised)
 DT 18-NOV-1996 (first entry)
 XX Carbonyl reductase coding sequence.
 DE carbonyl reductase; R-gamma-substituted-beta-hydroxybutyrate;
 KW gamma-substituted acetoacetate; raw material; drug synthesis;
 KW agricultural chemical production; ds.
 XX Sporidiobolus salmonicolor; IF01038.
 OS
 XX Key Location/Qualifiers
 FH 1. .972
 FT /tag= a
 FT /product= "carbonyl_reductase"
 FT 4. .969
 FT mat_peptide
 FT /tag= b
 XX JP08103269-A.
 XX 23-APR-1996.
 PD 07-OCT-1994; 94JP-00244090.
 XX 07-OCT-1994; 94JP-00244090.
 PR (ELED) DENKI KAGAKU KOGYO KK.
 XX WPI; 1996-253873/26.
 DR P-PSDB; AAR96294.
 XX Carbonyl reductase gene and protein - used in the production of R-gamma-
 FT substituted-beta-hydroxybutyrate, for use in drugs and agrochemicals.
 XX Claim 3; Page 10-11; 14pp; Japanese.
 XX The present sequence encodes a carbonyl reductase (ALD) isolated from
 CC Sporidiobolus salmonicolor IF01038. The enzyme is used in the synthesis
 CC of an (R)-gamma-substituted-beta-hydroxybutyrate (GSBH) from a gamma-
 CC substituted acetoacetate. The GSBH is useful as a raw material for the
 CC synthesis of drugs and agricultural chemicals. (Updated on 16-OCT-2003 to
 CC standardise OS field)
 XX Sequence 972 BP; 203 A; 337 C; 265 G; 167 T; 0 U; 0 Other;
 SQ
 Query Match 20.4%; Score 199.6; DB 2; Length 972;
 Best Local Similarity 55.5%; Pred. No. 2.5e-43;
 Matches 463; Conservative 0; Mismatches 344; Indels 27; Gaps 3;
 QY 87 GGGCGAGACCTATCTACTGTCTGTCACCTGCGCCCTGAAGACCGGTTACCGTCACCTTGCACTG 146
 DB 75 GGGCGAGGTGGCCAGGGCGTCAAGTTCGCCATCGAGACTGGATACCGTCACCTCGACCT 134
 QY 147 TGCTGTGTACTACTGAACGAGGGTGAGTTGTGAGGGTATCCGTGACTTCCTGAAGGA 206
 DB 135 TGCCAAAGGTCTACTCGAACCAACCTGAGTTGTGCGCCATCA-----A 179
 QY 207 GAACCCCTCGGTGAAGCGTGAGGACATCTTCGTCTGCACCAAGGTGGAAACCACTCCA 266
 DB 180 GGAGGTGGCGTCAAGCGGAGGACCTCTTCATCATCCTCGAGCTCTGGAAACACTCGCA 239
 QY 267 CCGTTATGAGGACCTCTCTGTGTCATTGACGACTCCCTGAAGCGTCTTGGAATTGACTA 326
 DB 240 CCGCCCGGAGCGGTGCGAGCTGCGCTTGAACACACCTCAAGGAGCTCGCCCTCGAGTA 299
 QY 327 CGTTGATATCT 386
 DB 300 CCTCGACCTTTACT 359
 QY 387 CAAGATTGGCCCTGACCGCAATAACGTCTTCTCAAG---GACCTGACCGAGAACCCCGA 443

DB 360 GAACCTCTTCCGAAGGCCAAGCAAGAGAGGTCAAGCTCGACTCGAGGTCAAGCTCGT 419
 QY 444 GCCCATATGGCGCGCTATGGAGAAGATTATGAGATCGCAAGGCCAGGTCCATTGGTGT 503
 DB 420 CGACACGTGGGAAGCGGATGGTCAAGCTTCTCGACACTGGCAAGGTCAAGCGGATCGGCT 479
 QY 504 CTCCAACTGGACCATTCGCGACCTTGAGAAGATGTCGAAGTTCCGAAGTCAATGCCCTCA 563
 DB 480 TTCCAACATTCGACGCGAAGATGGTCGAGCCCATCATCGAGGCTACCGCGGTGACCCCTC 539
 QY 564 CGCCCAACCAAGATCGAGATTACACCCCTTCTCTGCCCAACGAGGAGCTGTGTCAGTACTGTT 623
 DB 540 CGTCAACCAAGATCGAGCGTCAACCTCTCTTCTCCAGCCGAGGTCTATCGCCACCAAA 599
 QY 624 CTCGAAGAAATATGCCCCGTGGCTACTCTCTCTGCCCTCGAGAACCAAGGTTCAC 683
 DB 600 GGCCAAGAACATTCACATTAACCGCATACTCTCTCTCGGTAAACACACCGTCGCGCGGCC 659
 QY 684 CACCGGTGAGCGGTCAGCGAGAAACAACTCTGAACGAGATCGCCGAGAGGCGGCA 743
 DB 660 TCTTCTTGTCCAGCACCCCGAGATCAAGGC-----ATCGCGAGAGAACGGCTG 710
 QY 744 CACCTTGTCTAGGTCTTATGTGCTGGGTCTCGCGCGTGGCTAGTCTCTCTCTCCCAA 803
 DB 711 CACGCCCGCTCAGGTCTCTCATTTGCTGGCCCATCGTTGGCGGCCACTCGGTATATCCCAA 770
 QY 804 GAGCTCCAAACCCCAAGCGCATTTGATGTCCTCAAGAGCATTTGAGCTCTCCGATCGGA 863
 DB 771 GTCGTCACCCCTCCCGCATTTGGGAGAACTTCAAGCAGTCTCGCTCTCGAGGAGA 830
 QY 864 CTTTGAAGCCATCAATGCGTTCGCAAGGTCGTCACTTCGTTTCGTTCAACAT 917
 DB 831 CGTCGATCCGCTCAGCAAGCTCGCGAGGTTGGGCGCGAGGGCTACAAACAT 884
 RESULT 12
 AAH74577
 ID AAH74577 standard; DNA; 972 BP.
 XX AC AAH74577;
 XX 11-SEP-2003 (revised)
 DT 15-OCT-2001 (first entry)
 XX Nucleotide sequence of a NADPH-dependent aldehyde reductase (ALR).
 XX Methionine gamma-lyase; mdeA gene; free folding energy; gene shuffling;
 KW directed evolution; molecular breeding;
 KW NADPH-dependent aldehyde reductase; ss.
 XX Sporidiobolus salmonicolor.
 OS Key Location/Qualifiers
 FH 1. .972
 FT /tag= a
 FT /transl_except= (pos: 316. .321, aa: Ile)
 FT /product= "NADPH-dependent aldehyde reductase (ALR)"
 XX WO200155342-A2.
 XX 02-AUG-2001.
 XX 31-JAN-2001; 2001WO-US003186.
 XX 31-JAN-2000; 2000US-00494921.
 PR 08-DEC-2000; 2000US-00734237.
 XX (BIOC-) BIOCATALYTICS INC.
 PA Rozzell DJ, Bui P, Hua L;
 XX WPI; 2001-483235/52.
 DR P-PSDB; AAG63561.

XX PD 12-JUN-2002.
XX PF 07-DEC-2001; 2001EP-00310251.
XX KW 07-DEC-2000; 2000JP-00372704.
XX KW 15-JAN-2001; 2001JP-00006144.
XX PR 02-FEB-2001; 2001JP-00026594.
XX PR 11-JUN-2001; 2001JP-00175175.
XX PA (SUMO) SUMITOMO CHEM CO LTD.
XX PI Asako H, Matsumura K, Shimizu M, Ito N, Wakita R;
XX DR WPI; 2002-550350/59.
XX PT New protein preferentially producing (S)-4-bromo-3-hydroxybutanoate by
XX PT asymmetrically reducing 4-bromo-3-oxobutanoate, useful in pharmaceuticals
XX PT and agrochemicals.
XX PS Example 1; Page 38; 56pp; English.
XX CC The present sequence represents a PCR amplified fragment of a gene
XX CC encoding a Penicillium citrinum protein. This protein is capable of
XX CC producing (S)-4-bromo-3-hydroxybutanoate by asymmetrically reducing 4-
XX CC bromo-3-oxobutanoate. The amplified fragment was inserted into plasmid
XX CC pBR-2. The protein and polynucleotides are useful for producing optically
XX CC active (S)-4-bromo-3-hydroxybutanoate, which is useful as an intermediate
XX CC in the production of pharmaceuticals and agrochemicals. The (S)-4-bromo-3-
XX CC -hydroxybutanoate produced can also be used to produce 4-cyano-3-
XX CC hydroxybutanoic acid
XX SQ Sequence 743 BP; 183 A; 202 C; 156 G; 167 T; 0 U; 35 Other;
XX
Query Match 42.1%; Score 412; DB 6; Length 743;
Best Local Similarity 100.0%; Pred. No. 1.9e-100;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 567 CAACGATCGAGATTACCCCTTCCTGCCCAACGAGGAGTGTGCGTACTGCTTCTC 626
Db 1 CAACGATCGAGATTACCCCTTCCTGCCCAACGAGGAGTGTGCGTACTGCTTCTC 60
QY 627 CAAGAACATTATGCGGTGGCTACTCTCTCTGGTTCGAGAACGAGTTCCACCAC 686
Db 61 CAAGAACATTATGCGGTGGCTACTCTCTCTGGTTCGAGAACGAGTTCCACCAC 120
QY 687 CGGTGAGCGGTTCAGCGAGAACAGACTCTCAACGAGATCGCGAGAGGCGCAACAC 746
Db 121 CGGTGAGCGGTTCAGCGAGAACAGACTCTCAACGAGATCGCGAGAGGCGCAACAC 180
QY 747 CTTGCTCAGGTTCTTATGCTGGGTCTGCGCGGTGCTACGCTTCTCCCAAGAG 806
Db 181 CTTGCTCAGGTTCTTATGCTGGGTCTGCGCGGTGCTACGCTTCTCCCAAGAG 240
QY 807 CTCACCCCAAGCGCATTTAGTCCACTTCAAGAGCATTTAGCTCTCCGATCCGACTT 866
Db 241 CTCACCCCAAGCGCATTTAGTCCACTTCAAGAGCATTTAGCTCTCCGATCCGACTT 300
QY 867 TGAAGCCATCAATGCCGTGTCAGAGGTGCTCACTTCGTTTCGTCAACATGAAGATAC 926
Db 301 TGAAGCCATCAATGCCGTGTCAGAGGTGCTCACTTCGTTTCGTCAACATGAAGATAC 360
QY 927 TTTCGGATATGATGCTGGCCGAGGAGACCGCCCAAGAACCTGTCGCTGA 978
Db 361 TTTCGGATATGATGCTGGCCGAGGAGACCGCCCAAGAACCTGTCGCTGA 412

RESULT 10
ABL59387
ID ABL59387 standard; DNA; 331 BP.
XX
AC ABL59387;
XX

DT 22-OCT-2002 (first entry)
XX DE PCR amplicon of a protein producing (S)-4-bromo-3-hydroxybutanoate.
XX KW (S)-4-bromo-3-hydroxybutanoate; 4-bromo-3-oxobutanoate; pharmaceutical;
XX KW agrochemical; 4-cyano-3-hydroxybutanoic acid; ss.
XX OS Penicillium citrinum.
XX PN EP1213354-A2.
XX PD 12-JUN-2002.
XX PF 07-DEC-2001; 2001EP-00310251.
XX PR 07-DEC-2000; 2000JP-00372704.
XX PR 15-JAN-2001; 2001JP-00006144.
XX PR 02-FEB-2001; 2001JP-00026594.
XX PR 11-JUN-2001; 2001JP-00175175.
XX PA (SUMO) SUMITOMO CHEM CO LTD.
XX PI Asako H, Matsumura K, Shimizu M, Ito N, Wakita R;
XX DR WPI; 2002-550350/59.
XX PT New protein preferentially producing (S)-4-bromo-3-hydroxybutanoate by
XX PT asymmetrically reducing 4-bromo-3-oxobutanoate, useful in pharmaceuticals
XX PT and agrochemicals.
XX PS Example 1; Page 38; 56pp; English.
XX CC The present sequence represents a PCR amplified fragment of a gene
XX CC encoding a Penicillium citrinum protein. This protein is capable of
XX CC producing (S)-4-bromo-3-hydroxybutanoate by asymmetrically reducing 4-
XX CC bromo-3-oxobutanoate. The amplified fragment was inserted into plasmid
XX CC pBR-1. The protein and polynucleotides are useful for producing optically
XX CC active (S)-4-bromo-3-hydroxybutanoate, which is useful as an intermediate
XX CC in the production of pharmaceuticals and agrochemicals. The (S)-4-bromo-3-
XX CC -hydroxybutanoate produced can also be used to produce 4-cyano-3-
XX CC hydroxybutanoic acid
XX SQ Sequence 331 BP; 70 A; 93 C; 92 G; 75 T; 0 U; 1 Other;
XX
Query Match 27.1%; Score 265; DB 6; Length 331;
Best Local Similarity 100.0%; Pred. No. 4.4e-61;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCTAACGGAAGACTTTCACATTGAGCAACGGCGTCAAGATTCCTGGCGTCGGCTTT 60
Db 67 ATGCTAACGGAAGACTTTCACATTGAGCAACGGCGTCAAGATTCCTGGCGTCGGCTTT 126
QY 61 GGTACCTTCGCTAGTGAAGTTTCCAAAGGCGAGACCTATCTGCTGCACCACTGCCCTG 120
Db 127 GGTACCTTCGCTAGTGAAGTTTCCAAAGGCGAGACCTATCTGCTGCACCACTGCCCTG 186
QY 121 AAGACGGTTACCGTCACTTGGACTGTGCTGGTACTACTCTGAACAGGGTGAGGTTGGT 180
Db 187 AAGACGGTTACCGTCACTTGGACTGTGCTGGTACTACTCTGAACAGGGTGAGGTTGGT 246
QY 181 GAGGTTATCCGTGACTTCTCTGAAGGAGAACCCCTCGGTGAAGCGTGAAGCATCTTCCTC 240
Db 247 GAGGTTATCCGTGACTTCTCTGAAGGAGAACCCCTCGGTGAAGCGTGAAGCATCTTCCTC 306
QY 241 TCACCAAGGTTGGAACCACTCC 265
Db 307 TGCACCAAGTGTGGAACCACTCC 331
RESULT 11
AAT29160
ID AAT29160 standard; cDNA to mRNA; 972 BP.
XX

Query Match	42.6%	Score 417;	DB 6;	Length 417;
Best Local Similarity	100.0%;	Pred. No. 6.9e-102;		
Matches 417;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
245	CCRAGGNTGGAACCACTCCACCGTTATGAGACGTCTCTGGTCCATTGACGACTCCC	304		
1	CCRAGGTTGGAACCACTCCACCGTTATGAGACGTCTCTGGTCCATTGACGACTCCC	60		
305	TGAAGCGCTCTTGGACTTTGACTACGTTTGATATGTTTCTCGTTCACTGGGCCATTGCTGCGC	364		
61	TGAAGCGCTCTTGGACTTTGACTACGTTTGATATGTTTCTCGTTCACTGGGCCATTGCTGCGC	120		
365	AGAAGAATGGCCAGGGTGAGCCCAAGATTGGCCCTGACGGCAAATACGTCATTCTCAAGG	424		
121	AGAAGAATGGCCAGGGTGAGCCCAAGATTGGCCCTGACGGCAAATACGTCATTCTCAAGG	180		
425	ACCTGACCGAGAACCCCGAGCCACATGCGCGCTATGGAGAAGATTTTATGAGGATCGCA	484		
181	ACCTGACCGAGAACCCCGAGCCACATGCGCGCTATGGAGAAGATTTTATGAGGATCGCA	240		
485	AGGCCAGGTGCATTGGTGTCTCAAATGAGACCAATTGCGCACTTGAGAAGATGTCCAAGT	544		
241	AGGCCAGGTGCATTGGTGTCTCAAATGAGACCAATTGCGCACTTGAGAAGATGTCCAAGT	300		
545	TCGCCAAGGTCATGCTTCAGCGCAAACAGATCGAGATTCAACCCCTTCCTGCCCAACGAGG	604		
301	TCGCCAAGGTCATGCTTCAGCGCAAACAGATCGAGATTCAACCCCTTCCTGCCCAACGAGG	360		
605	AGCTGGTGCAGTACTGCTTCTCAAAGAACATTATGCCGTGGCCCTACTCTCTCTGG	661		
361	AGCTGGTGCAGTACTGCTTCTCAAAGAACATTATGCCGTGGCCCTACTCTCTCTGG	417		

RESULT 8
ABZ53802
IID ABZ53802 standard; CDNA; 587 BP.
XX
XX AC ABZ53802;
XX
XX 28-MAR-2003 (first entry)
DT

XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX PPA (NARE-) NAT RES INST BREWING.
XX PPA (NORQ) NAT FOOD RES INST MIN AGRIC.
XX
XX Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
XX PPI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
XX
XX WPI; 2003-046817/04.
XX
XX Detection of expression of specific *Aspergillus* genes for monitoring the
XX fermentation and growth conditions of the fungus, using DNA probes.
XX
XX Claim 1; SEQ ID NO 2915; 48pp + Sequence Listing; Japanese.
XX
XX The invention relates to a polynucleotide having any of 6006 specific
XX sequences (AB250888-AB256893), which are expressed by a fungus under
XX specific culture conditions including one or more of eutrophic,
XX

CC	oligotrophic, solid, early germination, alkaline, high temperature, low
CC	temperature or maltose culture or polynucleotides stringently hybridising
CC	to these sequences. The polynucleotides are useful for monitoring the
CC	progress of fermentation and the growth conditions of a fungus,
CC	especially of <i>Aspergillus oryzae</i> which is widely used in industrial
CC	fermentation. Also monitoring for fungal contamination. Note: The
CC	sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 587 BP; 142 A; 177 C; 151 G; 117 T; 0 U; 0 Other;
	Query Match 42.5%; Score 415.4; DB 7; Length 587;
	Best Local Similarity 83.8%; Pred. No. 2.1e-101;
	Matches 470; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY	1 ATGCTAACGGAAGACTTTACATTGAGCAACGGCGTCAAGATTCTTGGCGTGGCGTTT 60
DB	27 ATGTGAAACGGAAGACTTTACCGCTGAGCAATGGCGTTAAGATCCCGGTGCGGTTTC 86
QY	61 GGTACCTTCGCTAGTGAAGGTTTCAAGGGCGAGACCTATACCTGTGTACACACTGCCCTG 120
DB	87 GGTACCTTCGCGAGTGAAGGCTTCAAGGGCGAGACCTACAGGCTGTCAACAAAGCCCTC 146
QY	121 AAGACCGGTTACCGTCACTTGGACTGTGCCTGTACTACTCAACAGAGGTGAGTTGGT 180
DB	147 GAGACCGGATACCGTCACTTGGACTGTGCCTGTGTTTACCTCAAGAGATGAGTTGGT 206
QY	181 GAGGATATCCGTGACTTCTCAAGAGAGAACCCCTCGGTGAAGCGTGAGGACATCTTCCTC 240
DB	207 GATGATATCCATGACTTCTCAAGAGAGAACCCCTCGGTCAAGCGGAGAGACATCTTCCTC 266
QY	241 TGCACCAAGGTGTGGAAACACTCTCAACGTTATAGGAGAGTCTCTGTGTCCATTGACGAC 300
DB	267 TGCACCTAAGTCTGGAATCACTTCCACCGTCCGAGGAGCTCCAGTGTGGTGGTGAATAC 326
QY	301 TCCCTGAAGCGCTTTGGACTTGACTACGTTTGATATGTTCTCGTTCACTGGGCCCATTGCT 360
DB	327 TCCCTGAAGAGACTCCGACTGGAATAGTTTGACTTCTTGTGTCCATGTGCCCATTGCC 386
QY	361 GCCGAGAGAAATGCCAGGGTGAGCCCAAGATTGGCCCTGACGGCAATACGTCAATTCTC 420
DB	387 TCCGAGAGAGGAGGACCAAGAGAAACCAAGATTGGCCCTGACGGCAAGTAGCTATCCTC 446
QY	421 AAGGACCTGACCGAGAACCCGAGCCCAACATGGCGCGCTATGGAAGATTTATGAGGAT 480
DB	447 AAGGAGCTCACTGAGAACCCCGAGCCCAACATGGCGCGCCATGGAAGATCTTACAGGAC 506
QY	481 CGCAAGGCGAGGTCATTTGGTGTCTCCAACTGGACCAATGCCGACTTTCGAGAAGATGTCC 540
DB	507 GGCAAGGCCAAGCCATCGTGTTTCCAACTGGACCAATCCCGGCTCGAGAAGCTGTTC 566
QY	541 AAGTTCGCCAAGTCAATGCTT 561
DB	567 AAGTTCGCCGAGATCAAGCCT 587

RESULT 9	
ABL59388	
ID	ABL59388 standard; DNA; 743 BP.
XX	
AC	ABL59388;
XX	
DT	22-OCT-2002 (first entry)
XX	
DE	PCR amplicon of a protein producing (S)-4-bromo-3-hydroxybutanoate.
XX	
KW	(S)-4-bromo-3-hydroxybutanoate; 4-bromo-3-oxobutanoate; pharmaceutical;
KW	agrochemical; 4-cyano-3-hydroxybutanoic acid; ss.
XX	
OS	Penicillium citrinum.
XX	
PN	EP1213354-A2.

XX 28-SEP-2000.
 XX 22-MAR-2000; 2000WO-US007791.
 XX 22-MAR-1999; 99US-00273623.
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 XX (NOVO) NOVO NORDISK AS.
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX WPI; 2000-594572/56.
 XX Monitoring differential expression of genes in filamentous fungal cells
 XX uses fluorescence-labeled nucleic acids isolated from the cells and a
 XX substrate of expressed sequence tags.
 XX Claim 98; Page 2986; 3161pp; English.
 XX The present invention describes a method for monitoring differential
 XX expression of genes in a first filamentous fungal (FF) cell relative to
 XX expression of the same genes in one or more second filamentous fungal
 XX cells. The method uses fluorescence-labeled nucleic acids isolated from
 XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 XX are used in the methods for monitoring differential expression of genes
 XX in a first filamentous fungal (FF) cell relative to expression of the
 XX same genes in one or more second filamentous fungal cells. Monitoring the
 XX global expression of genes from FF cells allows the production potential
 XX of the microorganisms to be improved. New genes may be discovered,
 XX possible functions of unknown open reading frames can be identified and
 XX gene copy number variation and stability can be monitored. The expression
 XX of genes can be used to study how FF cells adapt to changes in culture
 XX conditions, environmental stress, spore morphogenesis, recombination,
 XX metabolic or catabolic pathway engineering. Using ESTs provides several
 XX advantages over genomic or random cDNA clones including elimination of
 XX redundancy as one spot on an array equals one gene or open reading frame,
 XX and organisation of the microarrays based on function of the gene
 XX products to facilitate analysis of the results. AAF07478 to AAF11247
 XX represents ESTs from *Fusarium venenatum*; AAF11248 to AAF11853 represents
 XX ESTs from *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from
 XX *Aspergillus oryzae*; and AAF14879 to AAF15337 represents ESTs from
 XX *Trichoderma reesei*, which are all specifically claimed in the present
 XX invention.
 XX Sequence 637 BP; 162 A; 189 C; 154 G; 132 T; 0 U; 0 Other;
 Query Match 43.0%; Score 420.2; DB 3; Length 637;
 Best Local Similarity 82.0%; Pred. No. 1.e-102;
 Matches 496; Conservative 0; Mismatches 108; Indels 1; Gaps 1;
 QY 1 ATGTCTACGGAAGACTTTCACATTTGAGCAACGGCGTCAAGATCTCTGGCGTCGGCTTT 60
 DB 30 ATGTGGAACGGGAAGACTTTCACGCTGAGCAATGGCGTTAAGATCCCCGGTGTGGTTTC 89
 QY 61 GGTACCTTCGCTAGTGAAGGTTTCAAGGGCGAGACCTTACTGTCTGCACACTGCCCTG 120
 DB 90 GGTACCTTCGCGATGAGGGTCTCAAGGGCGAGACCTTACAGGCTGTACCAAGGCCCTC 149
 QY 121 AAGCCGGTTACCGTCACTTGGATGTGCTGCTGGTACTTACTGTAACGAGGGTGAAGTTGGT 180
 DB 150 GAGACCGGATACCGTCACTTGGATGTGCTGCTGGTACTTACTTCAACGAGGATGAAGTTGGT 209
 QY 181 GAGGTATCCGTGACTTCTCAAGGAGAACCCCTCGGTGAAGCGTGAAGCATCTTCGTC 240
 DB 210 GATGGTATCAATGACTTCTCAAGAGAACCCCTCGGTCAAGCGGGAAGACATCTTCGTC 269
 QY 241 TGCACCAAGGTGTGAACCACTCCACGGTATGAGGACGTCCTCTGGTCCATTAAGCAC 300
 DB 270 TGCACCAAGGTGTGAACCACTCCACGGTATGAGGACGTCCTCTGGTCCATTAAGCAC 329
 QY 301 TCCCTGAGCGCTTGGACTTGAATGATGCTTCTCTGTTCACTGGCCCATGCT 360

DB 330 TCCTGAGAAACTCGACTGCACTAGTTGACCTCTTCTTGTCTCACTGGCCCATGGCC 389
 QY 361 GCGGAGAAGATGGCCAGGGTGAGCCCAAGATTGGCCCTCGAGGCAAAATACGTCAATTCTC 420
 DB 390 TCCGAGAGGAGGACCCAGGAAAAACCCAAAGATTGGCCCTCGAGGCAAGTACGTCACTCTT 449
 QY 421 AAGGACCTGACCGAGAACCCCGAGCCCAATGCGCGCTATGGAGAAGATTATGAGGAT 480
 DB 450 AAGGAGCTCACTGAGAACCCCGAGCCCAATGCGCGCTATGGAGAAGATTATGAGGAT 509
 QY 481 CGCAAGGCCAGGTCCTCAATTGGTGTCTTCAACTGCGACCATTTGCCACCTTGGAGAAGATGTC 540
 DB 510 CGCAAGGCCCAAGCCATCGGTGT-TTCAACTGCGACCATTTCCGGGTCTCGAGAAGCTGTTTC 568
 QY 541 AAGTTCGCCCAAGGTCATGCTTCAAGCCCAACGATCGAGATTCACCCCTCTCTGCCCAAC 600
 DB 569 AAGTTCGGCGAGATCAAGCCCTTATGTTAACCAAGATGAGATCCACCCCTTTCTGGCCAAC 628
 QY 601 GAGGA 605
 DB 629 AAGA 633
 RESULT 7
 ID ABL59391 standard; DNA; 417 BP.
 XX ABL59391;
 XX 22-OCT-2002 (first entry)
 XX PCR amplicon of a protein producing (S)-4-bromo-3-hydroxybutanoate.
 XX (S)-4-bromo-3-hydroxybutanoate; 4-bromo-3-oxobutanoate; pharmaceutical;
 XX agrochemical; 4-cyano-3-hydroxybutanoic acid; ss.
 XX Penicillium citrinum.
 XX BP1213354-R2.
 XX 12-JUN-2002.
 XX 07-DEC-2001; 2001EP-00310251.
 XX 07-DEC-2000; 2000JP-00372704.
 XX 15-JAN-2001; 2001JP-00006144.
 XX 02-FEB-2001; 2001JP-00026594.
 XX 11-JUN-2001; 2001JP-00175175.
 XX (SUMO) SUMITOMO CHEM CO LTD.
 XX Asako H, Matsumura K, Shimizu M, Ito N, Wakita R;
 XX WPI; 2002-550350/59.
 XX New protein preferentially producing (S)-4-bromo-3-hydroxybutanoate by
 XX asymmetrically reducing 4-bromo-3-oxobutanoate, useful in pharmaceuticals
 XX and agrochemicals.
 XX Example 1; Page 39; 56pp; English.
 XX The present sequence represents a PCR amplified fragment of a gene
 XX encoding a penicillium citrinum protein. This protein is capable of
 XX producing (S)-4-bromo-3-hydroxybutanoate by asymmetrically reducing 4-
 XX bromo-3-oxobutanoate. The amplified fragment was inserted into plasmid
 XX pBR-3. The protein and polynucleotides are useful for producing optically
 XX active (S)-4-bromo-3-hydroxybutanoate, which is useful as an intermediate
 XX in the production of pharmaceuticals and agrochemicals. The (S)-4-bromo-3-
 XX hydroxybutanoate produced can also be used to produce 4-cyano-3-
 XX hydroxybutanoic acid
 XX Sequence 417 BP; 94 A; 126 C; 105 G; 92 T; 0 U; 0 Other;

Db 601 GAGGAGCTGGTGCAGTACTGCTTCTCCAGAAACATTATATGCCCCGCTACTCTCTCTG 660
Qy 661 GCCTCGCAGAACACGGTTCACACACCGGTGAGGGGTGACGAGACAGACTCTGAC 720
Db 661 GCCTCGCAGAACACGGTTCACACACCGGTGAGGGGTGACGAGACAGACTCTGAC 720
Qy 721 GAGATCCCGAGAAAGCGGCGCAACACCTTGGCTCAGGTTCTTATGCTGGGCTCTGCGC 780
Db 721 GAGATCCCGAGAAAGCGGCGCAACACCTTGGCTCAGGTTCTTATGCTGGGCTCTGCGC 780
Qy 781 CGTGGCTACGTGCTTCTCCCAAGAGCTCCAAACCCCAAGCGCATGAGTCCCACTTCAAG 840
Db 781 CGTGGCTACGTGCTTCTCCCAAGAGCTCCAAACCCCAAGCGCATGAGTCCCACTTCAAG 840
Qy 841 AGCATTCAGCTCTCCGATGCGGACTTTGAAGCCATCAATGCCGTGTCGAAGGCTCGTCA 900
Db 841 AGCATTCAGCTCTCCGATGCGGACTTTGAAGCCATCAATGCCGTGTCGAAGGCTCGTCA 900
Qy 901 TTCGGTTTCGTCACATGAAGGATACCTTCGGATATGATCTGCGCCGAGGAGACCGCC 960
Db 901 TTCGGTTTCGTCACATGAAGGATACCTTCGGATATGATCTGCGCCGAGGAGACCGCC 960
Qy 961 AAGAACCTGCTCGGTGA 978
Db 961 AAGAACCTGCTCGGTGA 978

RESULT 5

ABL59384

ID ABL59384 standard; DNA; 697 BP.

XX

AC ABL59384;

XX

DT 22-OCT-2002 (first entry)

XX

DE PCR amplicon of a protein producing (S)-4-bromo-3-hydroxybutanoate.

XX

XX (S)-4-bromo-3-hydroxybutanoate; 4-bromo-3-oxobutanoate; pharmaceutical;

KW

KW agrochemical; 4-cyano-3-hydroxybutanoic acid; ss.

XX

XX Penicillium citrinum.

XX

PN EP1213354-A2.

XX

PD 12-JUN-2002.

XX

PF 07-DEC-2001; 2001EP-00310251.

XX

XX 07-DEC-2000; 2000JP-00372704.

PR

PR 15-JAN-2001; 2001JP-00006144.

PR

PR 02-FEB-2001; 2001JP-00026594.

PR

PR 11-FEB-2001; 2001JP-00175175.

XX

XX (SUMO) SUMITOMO CHEM CO LTD.

XX

XX Asako H, Matsumura K, Shimizu M, Ito N, Wakita R;

PI

XX WPI; 2002-550350/59.

XX

XX New protein preferentially producing (S)-4-bromo-3-hydroxybutanoate by

XX

XX asymmetrically reducing 4-bromo-3-oxobutanoate, useful in pharmaceuticals

XX

XX and agrochemicals.

XX

XX Example 1; Page 37; 56pp; English.

XX

XX The present sequence represents a PCR amplified fragment of a gene

CC

CC encoding a Penicillium citrinum protein. This protein is capable of

CC

CC producing (S)-4-bromo-3-hydroxybutanoate by asymmetrically reducing 4-

CC

CC bromo-3-oxobutanoate. The amplified fragment was inserted into plasmid

CC -hydroxybutanoate produced can also be used to produce 4-cyano-3-
XX hydroxybutanoic acid

SQ Sequence 697 BP; 161 A; 194 C; 172 G; 153 T; 0 U; 17 Other;

Query Match 60.2%; Score 588.6; DB 6; Length 697;

Best Local Similarity 94.4%; Pred. No. 6.5e-148;

Matches 611; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

Qy 4 TCTAACGGAAGACATTTTCACATTTGAGCAACGCGGTCAAGATTCCTGGCGTGGCTTTGGT 63

Db 52 TCCACGGAANACTTTTCACATTTGAGCAACGCGGTCAAGATTCCTGGCGTGGCTTTGGT 111

Qy 64 ACCTTCGCTAGTGAAGGTTCCAGGCGAGACCTTACTGTCTGTCTGTCACACTGCCCTGAAG 123

Db 112 ACCTTCGCTAGTGAAGGTTCCAGGCGGAAACCTTACTGTCTGTCTGTCACACTGCCCTGAAG 171

Qy 124 ACCGGTTACCGTCACTTTGAGCTGTGCTTGTACTTACCTGAAACGAGGGTGAAGTTGGTGAG 183

Db 172 ACCGGTTACCGTCACTTTGAGCTGTGCTTGTACTTACCTGAAACGAGGGTGAAGTTGGTGAG 231

Qy 184 GGTATCCGTGACTTCTCAAGAGAACCCCTCGGTGAAGCGTGAAGCATCTTGTCTGTC 243

Db 232 GGTATCCGTGACTTCTCAAGAGAACCCCTCGGTGAAGCGTGAAGCATCTTGTCTGTC 291

Qy 244 ACCAAGGTTGGAACCACTCCACCGTTATGAGGACGTCCTCTGCTGTCATTGACGACTCC 303

Db 292 ACCAAGGTTGGAACCACTCCACCGTTATGAGGACGTCCTCTGCTGTCATTGACGACTCC 351

Qy 304 CTGAAGCGTCTTGGACTTTGACTTACCTTGCATATGTTCTCTGCTCACTGCGCCATTGCTGCC 363

Db 352 CTGAAGCGTCTTGGACTTTGACTTACCTTGCATATGTTCTCTGCTCACTGCGCCATTGCTGCC 411

Qy 364 GAGAAGATGGCCAGGTTGAGCCGAGATTTGGCTGTGAGCGCAATACGTCTCTCAAG 423

Db 412 GAAAAAATGGCCAGGTTGAGCCGAGATTTGGCTGTGAGCGCAATACGTCTCTCAAG 471

Qy 424 GACCTGACCGAGAACCCGAGCCACATGGCGCGCTATGAGGAGAGATTTATGAGGATCGC 483

Db 472 GACCTGACCGA-AANCCNANCCACCTGGCGCGCTATGAAAAAATTTTNGANGATCCC 530

Qy 484 AAGCGCAGTCCATTGGTGTCTCCAACTGGACCATTCGGACCTTGAGAAAGATGTCGAAG 543

Db 531 AAGCGCAGTCCATTGGTGTCTCCAACTGGACCATTCGGACCTTGAGAAAGATGTCGAAG 590

Qy 544 TTGCGCCAGGTCATGCTTCCAGCCAAACAGATTCAGAGATTCACCCCTTCTGCCCCAACGAG 603

Db 591 TTGCGCCAGGTCATGCTTCCAGCCAAACAGATTCAGAGATTCACCCCTTCTGCCCCAACGAG 650

Qy 604 GAGCTGGTGCAGTACTGCTTCTCCAGAAACATTTATGCGCGTGGCGCTA 650

Db 651 GAGCTGGTGCAGTACTGCTTCTCCAGAAACATTTATGCGCGTGGCGCTA 697

RESULT 6

AAF14590

ID AAF14590 standard; cDNA; 637 BP.

XX

AC AAF14590;

XX

DT 13-MAR-2001 (first entry)

XX

XX Aspergillus oryzae EST SEQ ID NO:7113.

XX

Multiple gene expression; filamentous fungal cell; EST;
Expressed sequence tag; Fusarium venenatum; Aspergillus niger;
Aspergillus oryzae; Trichoderma reesei; identification; recombination;
culture condition; environmental stress; spore morphogenesis;
metabolic pathway engineering; catabolic pathway engineering; ss.

Aspergillus oryzae.

OS

XX

XX WO200056762-A2.

CC useful for producing optically active (S)-4-bromo-3-hydroxybutanoate,
CC which is useful as an intermediate in the production of pharmaceuticals
CC and agrochemicals. The (S)-4-bromo-3-hydroxybutanoate produced can also
CC be used to produce 4-cyano-3-hydroxybutanoic acid
XX
SQ Sequence 978 BP; 218 A; 282 C; 261 G; 217 T; 0 U; 0 Other;

Query Match 100.0%; Score 978; DB 6; Length 978;

Best Local Similarity 100.0%; Pred. No. 1.7e-252;

Matches 978; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTAACGAAAGACTTTCACATTGAGCAACGGCGTCAAGATTCTCGCGTGGCTTT 60
DB 1 ATGCTAACGAAAGACTTTCACATTGAGCAACGGCGTCAAGATTCTCGCGTGGCTTT 60
QY 61 GGTACCTTCGTAGTGAAGTTCACAGGGCGAGACCTATATCTGTCTGTCACCACTGCCCTG 120
DB 61 GGTACCTTCGTAGTGAAGTTCACAGGGCGAGACCTATATCTGTCTGTCACCACTGCCCTG 120
QY 121 AAGACCGGTTACCGTCACTTGGACTGTCCCTGTACTACCTGAACGAGGGTGAGTTGGT 180
DB 121 AAGACCGGTTACCGTCACTTGGACTGTCCCTGTACTACCTGAACGAGGGTGAGTTGGT 180
QY 181 GAGGTATCCGTGACTTCTTCAAGGAGAACCCCTCGGTGAAGCGTGAGGACATCTTCGTC 240
DB 181 GAGGTATCCGTGACTTCTTCAAGGAGAACCCCTCGGTGAAGCGTGAGGACATCTTCGTC 240
QY 241 TGCACCAAGGTGTGAACCACTCTCAGCTTATGAGGACGTCCTCTGTCATTTGAGGAC 300
DB 241 TGCACCAAGGTGTGAACCACTCTCAGCTTATGAGGACGTCCTCTGTCATTTGAGGAC 300
QY 301 TCCCTGAAGCGTCTTGGACTTGAATGATGATGTTCTCGTTCACTGGCCCATTTGCT 360
DB 301 TCCCTGAAGCGTCTTGGACTTGAATGATGATGTTCTCGTTCACTGGCCCATTTGCT 360
QY 361 GCCGAGAGAAATGCGCAGGGTGAGCCCAAGATTGGCCCTGACGGCAAAATACGTCATTC 420
DB 361 GCCGAGAGAAATGCGCAGGGTGAGCCCAAGATTGGCCCTGACGGCAAAATACGTCATTC 420
QY 421 AAGGACCTGACCGAGAACCCGAGCCACATCGCGCTATGGAGAGATTATGAGGAT 480
DB 421 AAGGACCTGACCGAGAACCCGAGCCACATCGCGCTATGGAGAGATTATGAGGAT 480
QY 481 CGCAAGCCAGGTCCATTGGTGTCTTCAACTTGGACCAATTCGCCGCTTGAAGAGATGCC 540
DB 481 CGCAAGCCAGGTCCATTGGTGTCTTCAACTTGGACCAATTCGCCGCTTGAAGAGATGCC 540
QY 541 AAGTCCGCAAGGTCTATGCTCAGCCCAACAGATCGAGATTCAACCCCTTCCGCGCAAC 600
DB 541 AAGTCCGCAAGGTCTATGCTCAGCCCAACAGATCGAGATTCAACCCCTTCCGCGCAAC 600
QY 601 GAGGAGCTGTGCACTGTCTTCCCAAGAACATTATGCCGCTGCGCTACTCTCTCTG 660
DB 601 GAGGAGCTGTGCACTGTCTTCCCAAGAACATTATGCCGCTGCGCTACTCTCTCTG 660
QY 661 GCCTCGAGAACAGGTTCCACACCGGTGAGCGGTGACGAGAACAGACTCTGAAC 720
DB 661 GCCTCGAGAACAGGTTCCACACCGGTGAGCGGTGACGAGAACAGACTCTGAAC 720
QY 721 GAGATCGCCGAGAGGGCGGCAACACCTTCTCAGGTTCTTATTGCTGGGCTCTGCG 780
DB 721 GAGATCGCCGAGAGGGCGGCAACACCTTCTCAGGTTCTTATTGCTGGGCTCTGCG 780
QY 781 CGTGGCTACGTTCTTCCCAAGAGTCTCAACCCCAAGCGCATTTAGTCCAACTTCAAG 840
DB 781 CGTGGCTACGTTCTTCCCAAGAGTCTCAACCCCAAGCGCATTTAGTCCAACTTCAAG 840
QY 841 AGCATGAGCTCTCGATGCGGACTTTGAAGCCATCAATGCCGTTGCGAGGTCGTCAC 900
DB 841 AGCATGAGCTCTCGATGCGGACTTTGAAGCCATCAATGCCGTTGCGAGGTCGTCAC 900
QY 901 TTCGGTTTCGTCACATGAAGGATCTTTGGATPATGATGTGCGCCCGAGGAGCCGCC 960
DB 901 TTCGGTTTCGTCACATGAAGGATCTTTGGATPATGATGTGCGCCCGAGGAGCCGCC 960

DB 901 TTCGGTTTCGTCACATGAAGGATCTTTGGATPATGATGTGCGCCCGAGGAGCCGCC 960
QY 961 AAGAACCTCTCTCGGTGA 978
DB 961 AAGAACCTCTCTCGGTGA 978

RESULT 2

ABL59398

ID ABL59398 standard; DNA; 978 BP.

XX ABL59398;

XX AC

XX XX

DT 22-OCT-2002 (first entry)

XX DNA sequence of a protein producing (S)-4-bromo-3-hydroxybutanoate.

XX (S)-4-bromo-3-hydroxybutanoate; 4-bromo-3-oxobutanoate; pharmaceutical;

KW agrochemical; 4-cyano-3-hydroxybutanoic acid; gene; ss.

XX OS

XX Penicillium citrinum.

XX FH

XX Key Location/Qualifiers

FT 1..978

FT /*tag= a

XX PN

EP1213354-A2.

XX XX

PD 12-JUN-2002.

XX XX

XX 07-DEC-2001; 2001EP-00310251.

XX XX

XX 07-DEC-2000; 2000JP-00372704.

PR 15-JAN-2001; 2001JP-00006144.

PR 02-FEB-2001; 2001JP-00026594.

PR 11-JUN-2001; 2001JP-00175175.

XX XX

XX (SUMO) SUMITOMO CHEM CO LTD.

XX XX

PI Asako H, Matsumura K, Shimizu M, Ito N, Wakita R;

XX WPI; 2002-550350/59.

DR P-PSDB; ABE77965.

XX XX

PT New protein preferentially producing (S)-4-bromo-3-hydroxybutanoate by
PT asymmetrically reducing 4-bromo-3-oxobutanoate, useful in pharmaceuticals
PT and agrochemicals.

XX PS

XX Example 5; Page 45-47; 56pp; English.

XX CC

CC The present sequence encodes a Penicillium citrinum protein which is
CC capable of producing (S)-4-bromo-3-hydroxybutanoate by asymmetrically
CC reducing 4-bromo-3-oxobutanoate. The protein and polynucleotides are
CC useful for producing optically active (S)-4-bromo-3-hydroxybutanoate,
CC which is useful as an intermediate in the production of pharmaceuticals
CC and agrochemicals. The (S)-4-bromo-3-hydroxybutanoate produced can also
CC be used to produce 4-cyano-3-hydroxybutanoic acid

XX SQ

Sequence 978 BP; 218 A; 282 C; 261 G; 217 T; 0 U; 0 Other;

XX Query Match 100.0%; Score 978; DB 6; Length 978;

XX Best Local Similarity 100.0%; Pred. No. 1.7e-252;

XX Matches 978; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTAACGAAAGACTTTCACATTGAGCAACGGCGTCAAGATTCTCGCGTGGCTTT 60

DB 1 ATGCTAACGAAAGACTTTCACATTGAGCAACGGCGTCAAGATTCTCGCGTGGCTTT 60

QY 61 GGTACCTTCGTAGTGAAGTTCACAGGGCGAGACCTATATCTGTCTGTCACCACTGCCCTG 120

DB 61 GGTACCTTCGTAGTGAAGTTCACAGGGCGAGACCTATATCTGTCTGTCACCACTGCCCTG 120

QY 121 AAGACCGGTTACCGTCACTTGGACTGTCCCTGTACTACCTGAACGAGGGTGAGTTGGT 180

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2004, 09:47:51 ; Search time 446 Seconds
(without alignments)
9315.555 Million cell updates/sec

Title: US-10-004-115B-2

Perfect score: 978
Sequence: 1 atgtctaacgaaagacttt.....ccaagaacctgtctgctga 978

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_29Jan04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002s:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	978	100.0	978	6	ABL59376
2	978	100.0	978	6	ABL59398
3	978	100.0	978	9	Ades39630 Penicilli
4	978	100.0	996	6	ABL59397 DNA seque
5	588.6	60.2	697	6	ABL59384 PCR ampli
6	420.2	43.0	637	3	AAFL4590 Aspergill
7	417	42.6	417	6	ABL59391 PCR ampli
8	415.4	42.5	587	7	ABZ53802 Aspergill
9	412	42.1	743	6	ABL59388 PCR ampli
10	265	27.1	331	6	ABL59387 PCR ampli
11	199.6	20.4	972	2	AAFL2160 Carboxyl
12	199.6	20.4	1055	2	AAFL2159 Carboxyl
13	195.6	20.4	978	4	ABL10491 Drosophil
14	178.8	18.3	978	4	ABZ56044 Aspergill
15	176.4	18.0	403	7	ABQ82227 T. megach
16	144.2	14.7	1121	6	ABQ82226 T. megach
17	142.6	14.6	1077	6	ABZ53102 Aspergill
18	135.4	13.8	656	7	ABZ53102 Aspergill
19	131.8	13.5	628	3	AAFL07918 Fusarium
20	129.4	13.2	1385	3	AAFL48070 Zea mays
21	127.2	13.0	963	4	ABL16915 Drosophil
22	120.8	12.4	1394	2	AAQ05879 Placenta-
23	119.2	12.2	1367	6	ABK09795 Human ova

24	119.2	12.2	1367	7	ABZ71966
25	119.2	12.2	1368	6	ABL58970
26	119	12.2	951	6	ABL01106
27	118	12.1	1292	4	ABL05161
28	117	12.0	1076	4	ABL16913
29	116.2	11.9	1059	7	ADA70389
30	116	11.9	1371	2	AAQ14946
31	115.2	11.8	1380	5	AAZ5873
32	113.2	11.6	942	5	AAH74584
33	113	11.6	1337	2	AAV58414
34	113	11.6	1337	8	ACF25340
35	113	11.6	1339	7	ABT41790
36	112	11.5	1163	3	AAZ50286
37	112	11.5	1163	3	AAZ45937
38	111.8	11.4	1913	3	AAZ77653
39	111	11.3	972	8	AAZ57091
40	111	11.3	972	8	AAZ57086
41	111	11.3	972	8	AAZ57088
42	111	11.3	972	9	AAZ42952
43	111	11.2	1290	8	AAZ57083
44	109.4	11.2	1064	8	AAZ57087
45	109.4	11.2	1165	3	AAZ33951

ALIGNMENTS

RESULT 1

ABL59376
ID ABL59376 standard; DNA; 978 BP.

XX ABL59376;

XX 22-OCT-2002 (first entry)

XX DE DNA sequence of a protein producing (S)-4-bromo-3-hydroxybutanoate.

XX KW (S)-4-bromo-3-hydroxybutanoate; 4-bromo-3-oxobutanoate; pharmaceutical;

XX KW agrochemical; 4-cyano-3-hydroxybutanoic acid; gene; ss.

XX OS Penicillium citrinum.

XX FH Key Location/Qualifiers

XX CDS 1..978

XX FT /*tag= a

XX PN EP1213354-A2.

XX PD 12-JUN-2002.

XX PF 07-DEC-2001; 2001EP-00310251.

XX PR 07-DEC-2000; 2000JP-00372704.

XX PR 15-JAN-2001; 2001JP-00006144.

XX PR 02-FEB-2001; 2001JP-00026594.

XX PR 11-JUN-2001; 2001JP-00175175.

XX PA (SUMO) SUMITOMO CHEM CO LTD.

XX PI Asako H, Matsumura K, Shimizu M, Ito N, Wakita R;

XX DR WPI; 2002-550350/59.

XX DR P-PSDB; ABB77965.

XX PT New protein preferentially producing (S)-4-bromo-3-hydroxybutanoate by

XX PT asymmetrically reducing 4-bromo-3-oxobutanoate, useful in pharmaceuticals

XX PT and agrochemicals.

XX PS Claim 1; Page 31-33; 56pp; English.

XX CC The present sequence encodes a Penicillium citrinum protein which is

XX CC capable of producing (S)-4-bromo-3-hydroxybutanoate by asymmetrically

XX CC reducing 4-bromo-3-oxobutanoate. The protein and polynucleotides are

Qy	362	CCGAGAGAAATGCGAGGTTAGCCCAAGATTGGCCCTGACGCAATACGTCATTCTCA	421
Db	245	CGGAGAGAAACCGAGCTATGATCAAGTCAAGATCGAGATGGAAGTAGATCATCAACA	304
Qy	422	AGGACCTGACCGAGAACCCGAGCCACATGGCGCGCTATGGAGAAAGATTATGAGGATC	481
Db	305	AAGAAGTCACGGCCAACTTGAACCTATTGGCGTAGATTTCGAGGCTCTTAACAAAGCTG	364
Qy	482	GCAAGGCCAGGTCGATTGGTGTCTCCAACTGGAGCAATTGCCGACCTTGAGAAGATGTCCA	541
Db	365	GCAAGGCAAAAGCCATCGAGTATCCAAATTTTACCATCTCAAACTGGAAGCTCTTCTCA	424
Qy	542	AGTTGCGCAAGGTCATGCCTCAGCCCAACAGATCGAGATTCACCCCTTCCTGCCCAACG	601
Db	425	AGTAGCGAGATGTTCTCCCGCCATCAACCAAGTCGAGATTCATCCGCTGSCCTAAACA	484
Qy	602	AGGAGCTGGTGCAGTACTGCTTCTCCAAAGAACATTATGCCGCTGGGCTACTCTCTCTGG	661
Db	485	CTAAGTTGATCAACTACTGTTTTTCAAGAACATCTCTTCGTTGGCATATTCGCCACTGG	544
Qy	662	GCTCCGAGAACCCAGGTTCCCAACACCGGTGAGCGGTCACGAGAACAGACTCTGAACG	721
Db	545	GATCCCGAGCCAGTCCCAACGACGAGGCAAAACCGTGATCCAGAACTCTGAACCTTATCT	604
Qy	722	AGATCGCCGAGAGGGCGGCAACACCCCTTGCTCAGGTTCTTATTGGCTGGGGTCTGCGCC	781
Db	605	TAATTCGCGAAAGAAAGGGCGTTAGCATAGTCAAAATATTGATTGTTGGTATCAAAAC	664
Qy	782	G-TGGCTACGTCGTTCTCCCGCAGAGCTCCAAACCCCAAGCGCATTGAGTCCAACTTCAAG	840
Db	665	GAAGGATACCTTAATTCCTTCCCATGAGTCAAAATGAGGGGAGATCAAAACCCCAATGGAACA	724
Qy	841	AGCAATTGAGCTCTCCGATGCCGACTTTGAAGCCATCAATGCCGTTGCCAAG	891
Db	725	CTTGTCGATTGACTGCAGAGGGTTCCGAAAATGACCCCAAGTTCGGAGG	775

Search completed: May 29, 2004, 12:17:05
Job time : 3881 secs

Qy 620 GCTTCTCAAGACATTTAGCCCTGCTACTCTCC 656
|||
Db 684 GCCTTCGAAGATTTGTACCACTTGCATATCCC 720
|||
RESULT 14
CNS01BYU 660 bp mRNA linear PLN 02-SEP-1999
LOCUS Botrytis cinerea strain T4 CDNA library under conditions of
DEFINITION nitrogen deprivation.
ACCESSION AL114862
VERSION AL114862.1 GI:5829481
KEYWORDS cdna library; nitrogen deprivation.
SOURCE Botryotinia fuckeliana
ORGANISM Botryotinia fuckeliana
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;
REFERENCE 1 (bases 1 to 660)
AUTHORS Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
78026 Versailles, France
REFERENCE 2 (bases 1 to 660)
GENOSCOPE.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT The cdna library to be analyzed within the framework of this
project was created using a Botrytis cinerea strain which was grown
under conditions of nitrogen deprivation, which is the normal
situation for B. cinerea during its development on its host plant.
The library was produced in an oriented direction, in the pBSII
vector.
FEATURES Location/Qualifiers
source 1..660
/organism="Botryotinia fuckeliana"
/mol_type="mRNA"
/strain="T4"
/db_xref="taxon:40559"
/note="Genoscope sequence ID : W71D061"
ORIGIN
Query Match 23.6%; Score 230.6; DB 8; Length 660;
Best Local Similarity 62.3%; Pred. No. 4.1e-39;
Matches 359; Conservative 0; Mismatches 217; Indels 0; Gaps 0;
Qy 20 TCACATTTAGCAACGGGTCAAGATTCTGCGCTGGCTTGTACCTTCGCTAGTGAAG 79
|||
Db 85 TCAAGCTGAACAATGGAGTTCAGATGCCAGCTTTTGGTTTCGGCACATTTGCCAGTGAGG 144
|||
Qy 80 GTTCCAGGGCGAGACCTTATCTGCTGCACCTGCCCTGAAGACCGGTTACCGTCACT 139
|||
Db 145 GTACGGTAGGAAGACACACACAGCCGCTGTAGCAGTTTGAACGGANGATACAGACTT 204
|||
Qy 140 TGGACTGTGCTGTACTCTGAACGAGGTGAGTTGGTGGGGTATCCGTGACTTCC 199
|||
Db 205 TGGACTGTGCTGTCTTACCAAAATGAGAACGAAGTCGGAACACCGCTTAAGGAGTTCC 264
|||
Qy 200 TGAAGGAGACCCCTCGTGAAGCGTGAAGGACATCTTCGCTGACCAACGAGGTGGAACC 259
|||
Db 265 TTGCTGCCAACCCCAAGTGTGAAGAGATCCGACATTTTCATGTGCACAAAGGTCTGGAATC 324
|||
Qy 260 ACCTCCACCGTTATGAGGACGCTCTCTGGTCCATTGACGACTCCCTGAAGCGTCTTGAC 319
|||
Db 325 ACCTCCACGACCTTGAAGATGCGAGTGGAGCTTGAAGAACTCGTTAGAAAGCTTCAAA 384
|||
Qy 320 TTGACTAGTTGATATGTTCTCGTTCACTGGCCCATTTGTCGCGAAGAGATGGCCAGG 379
|||
Db 385 CCCCTTACATTTGATGTCATTTGTTGACACTGGCCAAATGCTGCANAGAAGATGAGACA 444
|||
Qy 380 GTGAGCCCAAGATGGCCCTGACCGCAAAATACGTCATTTCTCAAGGACCTGACCGAGAACC 439
|||

Db 445 GATCAGTCAAGATCGGTGCTGATGGAAGTAGTCTCATCAAGAAAGATTTGACAGAAACC 504
|||
Qy 440 CCGAGCCACATGGCGCTATGGAGAAGATTATATGAGATCGCAAGCCAGGTCCATTG 499
|||
Db 505 CAGAGCCACATGGAGAGCATGGANNAAGTTATCAAGAGGGGACTCCGGAAGGTATCG 564
|||
Qy 500 GTGTCTCAACTGGACCATTCGCCACTTGGAGAGATCTCAAGTTCCGCAAGGTCTGCG 559
|||
Db 565 GTGTCTCAACTGGACCGAGGAAGTAATGAGCAGCTTTGTCTATCTCGTGAAGTTAAGC 624
|||
Qy 560 CTCACGCCAACCATCGAGATTCACCCCTTCCTGCG 595
|||
Db 625 CAACAATTAACCAATCGAGATCCATCTTTCCTTC 660
|||
RESULT 15
CNS01BQ2 780 bp mRNA linear PLN 02-SEP-1999
LOCUS Botrytis cinerea strain T4 CDNA library under conditions of
DEFINITION nitrogen deprivation.
ACCESSION AL114546
VERSION AL114546.1 GI:5829165
KEYWORDS cdna library; nitrogen deprivation.
SOURCE Botryotinia fuckeliana
ORGANISM Botryotinia fuckeliana
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;
REFERENCE 1 (bases 1 to 780)
AUTHORS Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
78026 Versailles, France
REFERENCE 2 (bases 1 to 780)
GENOSCOPE.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT The cdna library to be analyzed within the framework of this
project was created using a Botrytis cinerea strain which was grown
under conditions of nitrogen deprivation, which is the normal
situation for B. cinerea during its development on its host plant.
The library was produced in an oriented direction, in the pBSII
vector.
FEATURES Location/Qualifiers
source 1..780
/organism="Botryotinia fuckeliana"
/mol_type="mRNA"
/strain="T4"
/db_xref="taxon:40559"
/note="Genoscope sequence ID : W35C041"
ORIGIN
Query Match 23.1%; Score 226.2; DB 8; Length 780;
Best Local Similarity 57.6%; Pred. No. 3.5e-38;
Matches 444; Conservative 0; Mismatches 323; Indels 4; Gaps 2;
Qy 125 CCGGTTACCGTCACTTGGACTGTGCTGCTACTACTCTCAACGAGGCTGAGGTGTTGAGG 184
|||
Db 5 CAGGTTATCGCATCTAGATTGCGCTTGGTATTACAAGATGAAGAGAGTTGGTCTG 64
|||
Qy 185 GTATCCGTGACTTCTCTGAAGAGAACCCCTCGTGAACGGTGAAGACATCTTCGTTCTGA 244
|||
Db 65 GCTTGGGGAATCTTGTCCAGCAACCCCAAGGTCAAGCGTTCCGATATCTTCATCACC 124
|||
Qy 245 CCAAGGTGGGAACCACTCCACCGTT---ATGAGAGCTCTCTGGTCCATTTGACGACT 301
|||
Db 125 CGAAGTGTGGCTCATCTATCGGTTCTCTGAAGAGCTCGAGTGGAGTCTGATTACA 184
|||
Qy 302 CCTGAAGCGCTTGTGACTGTACTAGTTGATATGTTCTCGTTCACTGGCCCATTTGCTG 361
|||
Db 185 GTTAGAGAAACTAGGGTTCGATTATGTGGATTCTTTCTTGATGATTGGCCATTCGCGG 244
|||

COMMENT The cDNA library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal situation for B. cinerea during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.

FEATURES Location/Qualifiers
1..720
/organism="Botryotinia fuckeliana"
/mol_type="mRNA"
/strain="T4"
/db_xref="taxon:40559"
/note="Genoscope sequence ID : W26B011"

ORIGIN
Query Match 25.3%; Score 247; DB 8; Length 720;
Best Local Similarity 62.5%; Pred. No. 1.2e-42;
Matches 402; Conservative 0; Mismatches 240; Indels 1; Gaps 1;
20 TCACATTGAGCAAGCGGTCAAGATTCCTGGCGTGGCTTTGGTACCTTCGCTAGTGAAG 79
Db TCAAGCTGAACAATGGAGTTCAGATGCCGCTTTGGTTTCGGACATTTGCCAGTGAGG 129
Qy 80 GTTCCAAAGGGCGAGACCTATCTCTGTCCACACTGCGCTCAAGACCGGTACCGTCACT 139
Db GTAGCGTAGGAGACACACAGAGCCCTCGTAGCAGCTTTGAACGCGAGGATACAGACATT 189
Qy 140 TGGACTGTGCTGTACTACTCTGAACAGGGTGGTGGAGGGTATCCGTGACTTCC 199
Db TGGACTGTGCTGTGTTCTTACCAAAATGAGAACGAAAGTCGGAACGCGGTTAAGGAGTTC 249
Qy 200 TGAAGGAGAACCCCTCGGTGAAGGCTGAGGACATCTTCGTCTGCACCAAGCTGTGAACC 259
Db TTGCTGCCAACCAAGTGTGAGAGATCCGACATTTTCATGTGCACAAAGTCTGGAATC 309
Qy 260 ACCTCCACCGTATGAGGAGCTCTCTGTGTCATGTAGCAGCTCCCTGAAGCGTCTTGGAC 319
Db ACCTCCACGACCTGAAAGATGTCAGTGGAGCTTGAAGAACTCGTTAGAAAAGCTTCAAA 369
Qy 320 TTGACTAGTTGATATGTTCTCGTCTGCTACCTGGCCCATTTGTCGCGAGAAATGGCCAGG 379
Db CCCCTTACATTTGATGCTATCTTGATACCTGGCCCAATGCTGCAGAAAGATGAGGACAG 429
Qy 380 GTGAGCCCAAGATTGGCCCTGACCGCAAAATACGTCATTCTCAAGGACCTGACCGAGAAC 439
Db ATCAG-TCAGGATCGGTCTGATGTAAGTAGTGTATCAGGAAAGATTTCACAGAAAACC 488
Qy 440 CCGAGCCCAATGCGGCTATGAGGAGATTTATGAGGATCGCAAGGCCAGGTCCATTG 499
Db CAGAGCCCAACATGGAGAGCAATGGAGAAAGTTATACAGAGGAGGACTCGCGAAGGCTATCG 548
Qy 500 GTGTCTCAACTGACCATTCGCCACCTTCGAGAGATGTCCAAGTTCGCCCAAGGCTATGC 559
Db GTGTCTCAACTGACCGAGAGGAAATGAGAGCTTTTGTCTATTCGCTGAGGTTAAGC 608
Qy 560 CTCAGGCCCAACGAGATCGAGATTCACCCCTTCCTGCCCCAACGAGGAGCTGGTGCAGTACT 619
Db CAACAATTAACCAATCGAGATCCATCTTCTTCCACAAACAAAGCTCATCGACTACT 668
Qy 620 GCTTCTCCAGAACATTTATGCGGTGGCTACTCTCTCTCTGGG 662
Db GCCTTTTGAAGAGATTTGTACCAAGTTCGATATTCCTCCATTTGGG 711

RESULT 13
CNS019U1 Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation.
LOCUS AL112097
DEFINITION AL112097.1 Gi:5826716
ACCESSION cDNA library; nitrogen deprivation.
VERSION Botryotinia fuckeliana
KEYWORDS
SOURCE

ORGANISM Botryotinia fuckeliana
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE 1 (bases 1 to 720)
AUTHORS Bittou, F., Levis, C., Fortini, D., Pradier, J.M. and Brygoo, Y.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr, 78026 Versailles, France
2 (bases 1 to 720)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage : CP 5706 91057 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
JOURNAL - Web : www.genoscope.cns.fr)
COMMENT The cDNA library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal situation for B. cinerea during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.

FEATURES Location/Qualifiers
1..720
/organism="Botryotinia fuckeliana"
/mol_type="mRNA"
/strain="T4"
/db_xref="taxon:40559"
/note="Genoscope sequence ID : W11E111"

ORIGIN
Query Match 23.8%; Score 233.2; DB 8; Length 720;
Best Local Similarity 61.4%; Pred. No. 1.1e-39;
Matches 391; Conservative 0; Mismatches 244; Indels 2; Gaps 1;
Qy 20 TCACATTGAGCAAGCGGTCAAGATTCCTGGCGTGGCTTTGGTACCTTCGCTAGTGAAG 79
Db TCAAGCTGAACAATGGAGTTCAGATGCCGCTTTGGTTTCGGACATTTGCCAGTGAGG 145
Qy 80 GTTCCAAAGGGCGAGACCTATCTCTGTCAACACTGCGCTCAAGACCGGTACCGTCACT 139
Db GTAGCGTAGGAGACACACAGAGCCCTCGTAGCAGCTTTGAACGCGAGGATACAGACATT 205
Qy 140 TGGACTGTGCTGTACTACTCTGAACGAGGGTGGTGGAGGGTATCCGTGACTTCC 199
Db TGGACTGTGCTGTGTTCTTACCAAAATGAGAACGAAAGTCGGAACGCGGTAAAGAGTTC 265
Qy 200 TGAAGGAGAACCCCTCGGTGAAGGCTGAGGACATCTTCGTCTGCACCAAGGTGTGGAACC 259
Db TTGCTGCCAACCCAGTGTGAAGAGATCCGACATTTTCATTTGTCAAAAGCTTGAATC 325
Qy 260 ACCTCCACCGTATGAGGAGCTCTCTGTGTCATTTGAGACTCCCTGGAAGCGTCTTGGAC 319
Db ACCTCCACGACCTCGATGATGTCGAGTGGAGCTTGAAGAACTCGTTAGAAAAGCTTCAAA 385
Qy 320 TTGACTAGTTGATATGTTCTCGTCTGCTACCTGGCCCATTTGTCGCGAGAAATGGCCAGG 379
Db CCCCTTACATTTGATGCTATCTTGATACCTGGCCCAATGCTGCAGAAAGATGAGGAC 445
Qy 380 GTGAGCCCAAGATTGGCCCTGACCGCAAAATACGTCATTCTCAAGGACCTGACCGAGAAC 439
Db GATCAGTCAAGATCGGTGCTGATGGAAGTACGTCATCAAGAAAGATTTGACAGAAAACC 505
Qy 440 CCGAGCCCAATGCGGCTATGAGAGATTTATGAGGATCGCAAGGCCAGGTCCATTG 499
Db CAGAGCCCAACATGGAGAGCAATGGAGAAAG--TATCAAGAGGAGGACTCGCGAAGTATCG 563
Qy 500 GTGTCTCAACTGAGACCATTCGCCACCTTGAGAGATGTCGAAGTTCGCCAAGGTCTATGC 559
Db GTGTCTCAACTGGACCGGAGAGGAAATGAGCACTTTTGTCTATTCGTTGAGGTTAAGC 623
Qy 560 CTCAGGCCCAACGAGATCGAGATTCACCCCTTCCTGCCCAACGAGGAGCTGGTGCAGTACT 619
Db CAACAATTAACCAATTCGAGATCCATCTTCTTCCACANCAAAAGCTCATCGACTACT 683

```
Db 121 AGAAGATGCGCGGTGAGCCCAAGATTGGCCCTGACGGCAAAATACGTCATTCTCAAGG 180
Qy 425 ACCTGACCGAGAACCCCGAGCCACATGCGCGCTATGGAGAGATTATGAGATCGCA 484
Db 181 ACCTGACCGAGAACCCCGAGCCACATGCGCGCTATGGAGAGATTATGAGATCGCA 240
Qy 485 AGGCGAGGTCCATTGGTGTCTCCAACTGGACCATTTGGCGACCTTGGAGAAGATGTCGAAGT 544
Db 241 AGGCGAGGTCCATTGGTGTCTCCAACTGGACCATTTGGCGACCTTGGAGAAGATGTCGAAGT 300
Qy 545 TCGCAAGGTTCATGCTCAGCCCAACAGATTCGAGATTACCCCTTCTCCGCCAACGAGG 604
Db 301 TCGCAAGGTTCATGCTCAGCCCAACAGATTCGAGATTACCCCTTCTCCGCCAACGAGG 360
Qy 605 AGCTGGTGCAGTATGCTCTCCAAAGAACATTAATGCGCGGCTACTCTCCCTGG 661
Db 361 AGCTGGTGCAGTATGCTCTCCAAAGAACATTAATGCGCGGCTACTCTCCCTGG 417

RESULT 10
AX472804
LOCUS AX472804 743 bp DNA linear PAT 09-AUG-2002
DEFINITION Sequence 19 from Patent EP1213354.
ACCESSION AX472804
VERSION AX472804.1 GI:22207653
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Asako,H., Matsumura,K., Shimizu,M., Ito,N. and Wakita,R.
AUTHORS Process for producing optically active 4-halo-3-hydroxybutanoate
TITLE Patent: EP 1213354-A 19 12-JUN-2002;
JOURNAL Sumitomo Chemical Company, Limited (JP)
FEATURES
source
1. 743
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"

ORIGIN
Query Match 42.1%; Score 412; DB 6; Length 743;
Best Local Similarity 100.0%; Pred. No. 4.3e-78;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 567 CAACGAGATCGAGATTCACCCCTTCTGCGCCCAAGAGAGTGTGCGACTGCTCTC 626
Db 1 CAACGAGATCGAGATTCACCCCTTCTGCGCCCAAGAGAGTGTGCGACTGCTCTC 60

Qy 627 CAAGAACAATTATGCCGTGGCTACTCTCTCTGGGTTCGAGAACCCAGGTTCACAC 686
Db 61 CAAGAACAATTATGCCGTGGCTACTCTCTCTGGGTTCGAGAACCCAGGTTCACAC 120

Qy 687 CGGTGAGCGGTACGCGAGACAGACTCTCAACGAGATCGCCGAGAGCGGCGACAC 746
Db 121 CGGTGAGCGGTACGCGAGACAGACTCTCAACGAGATCGCCGAGAGCGGCGACAC 180

Qy 747 CTTGTCTCAGTCTTATTGCTGGGTCTCGCGCGGTGCTGCTGCTTCTCCCGAAG 806
Db 181 CTTGTCTCAGTCTTATTGCTGGGTCTCGCGCGGTGCTGCTGCTTCTCCCGAAG 240

Qy 807 CTCGAACCCCAAGCGCATTTAGTCAACTTCAAGAGATTGAGTCTCCGATCGCGACTT 866
Db 241 CTCGAACCCCAAGCGCATTTAGTCAACTTCAAGAGATTGAGTCTCCGATCGCGACTT 300

Qy 867 TGAAGCCATCAATGCGGTTCGCAAGGGTCTGCTCTTCTGTTCTCAACATGAAGGATAC 926
Db 301 TGAAGCCATCAATGCGGTTCGCAAGGGTCTGCTCTTCTGTTCTCAACATGAAGGATAC 360

Qy 927 TTTCGATATGATGTCTGCGCGCGAGGAGACCGCAAGAACCTGTCTCGTGA 978
Db 361 TTTCGATATGATGTCTGCGCGCGAGGAGACCGCAAGAACCTGTCTCGTGA 412
```

RESULT 11

```
AX472803
LOCUS AX472803 331 bp DNA linear PAT 09-AUG-2002
DEFINITION Sequence 18 from Patent EP1213354.
ACCESSION AX472803
VERSION AX472803.1 GI:22207652
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Asako,H., Matsumura,K., Shimizu,M., Ito,N. and Wakita,R.
AUTHORS Process for producing optically active 4-halo-3-hydroxybutanoate
TITLE Patent: EP 1213354-A 18 12-JUN-2002;
JOURNAL Sumitomo Chemical Company, Limited (JP)
FEATURES
Location/Qualifiers
1. 331
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"

ORIGIN
Query Match 27.1%; Score 265; DB 6; Length 331;
Best Local Similarity 100.0%; Pred. No. 1.8e-46;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTCTAACGGAAGACTTTCACATTGAGCAACGGCGTCAAGATTCTCTGGGTCGGCTTT 60
Db 67 ATGTCTAACGGAAGACTTTCACATTGAGCAACGGCGTCAAGATTCTCTGGGTCGGCTTT 126

Qy 61 GGTACCTTCGTAGTGAAGTTCCAAAGGCGAGACTTACTGCTCACCACCTGCCCTG 120
Db 127 GGTACCTTCGTAGTGAAGTTCCAAAGGCGAGACTTACTGCTCACCACCTGCCCTG 186

Qy 121 AAGACCGGTACCGTCACTTGGACTGTGCTGCTGCTACTACCTGAACGAGGGTGAGTTGGT 180
Db 187 AAGACCGGTACCGTCACTTGGACTGTGCTGCTGCTACTACCTGAACGAGGGTGAGTTGGT 246

Qy 181 GAGGGTATCCGTGACTTCTGAGGAGAACCCCTCGGTGAAGCGTGAGGACATCTTCGTC 240
Db 247 GAGGGTATCCGTGACTTCTGAGGAGAACCCCTCGGTGAAGCGTGAGGACATCTTCGTC 306

Qy 241 TGCACCAAGGTGTGGAACCACTCC 265
Db 307 TGCACCAAGGTGTGGAACCACTCC 331
```

RESULT 12

```
CNS01C4M
LOCUS CNS01C4M 720 bp mRNA linear PLN 02-SEP-1999
DEFINITION Botrytis cinerea strain T4 CDNA library under conditions of
nitrogen deprivation.
ACCESSION AL115070
VERSION AL115070.1 GI:5829689
KEYWORDS cDNA library; nitrogen deprivation.
SOURCE Botryotinia fuckeliana
ORGANISM Botryotinia fuckeliana
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE
1 (bases 1 to 720)
AUTHORS Bittori,F., Lewis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
78026 Versailles, France
REFERENCE
2 (bases 1 to 720)
Genoscope.
Direct Submission
Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
```

```

QY 604 GAGTGTGTCAGTACTGCTTCTCAAGAACATTATGCGCCGTGGCCTA 650
    |||||
Db 651 GAGCTGTGTCAGTACTGCTTCTCAAGAACATTATGCGCCGTGGCCTA 697
    |||||

RESULT 8
AB014493
LOCUS Gibberella zeae gene for reductase, partial cds. PLN 10-OCT-1998
DEFINITION Gibberella zeae
ACCESSION AB014493
VERSION AB014493.1 GI:3724333
KEYWORDS reductase.
SOURCE Gibberella zeae
ORGANISM Gibberella zeae
REFERENCE 1 (sites)
Eukaryote; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
AUTHORS Kimura, M., Matsumoto, G., Shingu, Y., Yoneyama, K. and Yamaguchi, I.
TITLE The mystery of the trichothecene 3-O-acetyltransferase gene.
ANALYSIS Analysis of the region around Trl101 and characterization of its
homologue from Fusarium sporotrichoides
JOURNAL FEBS Lett. 435 (2-3), 163-168 (1998)
MEDLINE 98433864
PUBMED 9762900
REFERENCE 2 (bases 1 to 762)
Kimura, M.
AUTHORS Direct Submission
TITLE Submitted (21-MAY-1998) Makoto Kimura, The Institute of Physical
and Chemical Research (RIKEN), Microbial Toxicology Laboratory;
Hirosewa 2-1, Wako, Saitama 351-0198, Japan
(E-mail: mkimura@postman.riken.go.jp, Tel: 81-48-467-9518,
Fax: 81-48-462-4676)
FEATURES
    Location/Qualifiers
        1..762
            /organism="Gibberella zeae"
            /mol_type="genomic DNA"
            /strain="F15"
            /db_xref="taxon:5518"
            <1..>762
            /note="putative"
            /codon_start=1
            /product="reductase"
            /protein_id="BA033773.1"
            /db_xref="GI:3724334"
            /translation="PGVGFQTFANEGAKGYKAVTAAKTYGRHLDCAWFYLNIEGV
GOAVRDLASDNGVKRDLFICTKWNHLHEPEVKSPENSLKNGFLDVIDLFLVHW
PIAAEKDDYQPKTPGDKYIKDLENPEPTWAMEEIVASGKARAIGVSNWTDG
LKQMAFAKVPANVQIEIHPFLPNEDLVKFLENDILPAAYSPFGSQNVPTGTIV
RSNKTLEAVDRSGNTLAQVLIANGURRGLFCFPA"
    ORIGIN
        Query Match 48.2%; Score 471; DB 8; Length 762;
        Best Local Similarity 76.9%; Pred. No. 9.2e-91;
        Matches 587; Conservative 0; Mismatches 175; Indels 1; Gaps 1;
QY 46 CCGCGGTCGCTTGTGTAACCTTCGCTAGTGAAGGTTCCAGGCGGAGACCTATCTGCT 105
Db 1 CCGCGGTCGCTTGTGTAACCTTCGCTAGTGAAGGTTCCAGGCGGAGACCTATCTGCT 60
QY 106 GTCAACCACTGCGCTGAAGACCGGTTACCGTACATGCTGCTGCTGCTGCTGCTGCTG 165
Db 61 GTCAATGCGGCTTCAGACCGGTTACCGTACATGCTGCTGCTGCTGCTGCTGCTGCTG 120
QY 166 GAGGTCGAGGTTGTGAGGTTATCCGTAACCTTCCTGAAGGAGAACCCCTCGGTGAACGCT 225
Db 121 GAGGTCGAGGTTGTGAGGTTATCCGTAACCTTCCTGAAGGAGAACCCCTCGGTGAACG 180
QY 226 GAGGACATCTTCGTCACCAAGTCTGGAACCACTCCACGCTTATGAGGACGCTCCTC 285
Db 181 GAGGATCTCTTCATTTGACAAAGGTTCTGGAACCACTCCACGCTTATGAGGAGGTTCA 240
QY 286 TGGTCCATTGACGACTCCCTGAAGCGCTCTTGGACTGACGTTGATATGTTCTCTCGTT 345

```

```

Db 241 TGGTCAATTCGAGAACTCTCTCAAGAACTTCGCGCTTGACTGACTTGTCTCTCGTC 300
QY 346 CACTGGCCCAATTGCTGCCGAGAAGAAATGCGCAGGGTGAGCCCAAGATTGGCCCTGAGCGC 405
Db 301 CACTGGCCCTATGCTGCCGAGAAGGGCGACGACTACCAAGCCCAAGATCGGCCCTGATGCG 360
QY 406 AATATAGCTCAITCTCAAGGACTTGACCGAGAACCCCGAGCCGACCATGATGCGGCTATGAG 465
Db 361 AAGTAGCTCATCAAGAAGACCTCACTGAGAACCCCGAGCCGACCATGCGGCGCATGAG 420
QY 466 AAGATTATGAGGATCGCAAGCCAGGTCCTATGTTGTCTCCAACTGGACCAATTCGCGAC 525
Db 421 GAGATTATGCTAGTGGCAAGGCCCGTCCCAITGTTGTCTCCAACTGGACCAATTCGCGAC 480
QY 526 CTTGAGAGATGTTCCAAAGTTGCGCAAGGTCATGCTCTACGGCAACCAAGATCGAGATTAC 585
Db 481 CTGAAGCAGCTCANGGCCCTTCGCAAGGTCAAGCCCGGTTAACCAAGATCGAGATTAT 540
QY 586 CCCTTCTCTGCCCAAGCAGGAGTGTGTCAGTACTGCTTCTCCAAAGCAATTCGCCCGTG 645
Db 541 CTTTCTCTGCCCAAGCAGGATCTGTCAGTCTGCTCGAGACGACATCTCTCCCGCT 600
QY 646 GCCTACTCTCTCTGCTGCGTCCGAGAACAGGTTCCCAACCGGTGAGCGGTCAGGAG 705
Db 601 GCTTACTCTCTCTGCTGCGTCCGAGAACAGGTTCCCAACCGGTGAGCGGTCAGGTC 660
QY 706 AACAGAGCTCTGAACGAGATCGCGAGAGGCGGCAACCGCTTGTCTGCTGCTTCTTATT 765
Db 661 AACAGAGCTCTGAACGAGATCGCGAGAGGCGGCAACCGCTTGTCTGCTGCTTCTTATT 720
QY 766 GCCTGGGTCGTCGCGCGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 808
Db 721 GCCTGGGTCGTCGCGCGGCGG-TAGCTGCTTCTGCGCAAGAGCT 762

RESULT 9
AX472807
LOCUS Sequence 22 from Patent EP1213354.
ACCESSION AX472807
VERSION AX472807.1 GI:22207656
KEYWORDS Escherichia coli
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE 1
AUTHORS Asako, H., Matsumura, K., Shimizu, M., Ito, N. and Wakita, R.
TITLE Process for producing optically active 4-halo-3-hydroxybutanoate
JOURNAL Patent: EP 1213354-A 22 JUN-2002;
Sumitomo Chemical Company, Limited (JP)
FEATURES
    source
        1..417
            /organism="Escherichia coli"
            /mol_type="unassigned DNA"
            /db_xref="taxon:562"
    ORIGIN
        Query Match 42.6%; Score 417; DB 6; Length 417;
        Best Local Similarity 100.0%; Pred. No. 3.8e-79;
        Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 245 CCAAGGTGTGGNACCACTCCACCGTTATGAGGAGCTCTCTGTCCTATTCAGGACTCCC 304
Db 1 CCAAGGTGTGGNACCACTCCACCGTTATGAGGAGCTCTCTGTCCTATTCAGGACTCCC 60
QY 305 TGAAGGCTCTTGGACTTGAAGTGTGATATGTTCTCTGTTCACTGGCCCAATTCGTGCG 364
Db 61 TGAAGGCTCTTGGACTTGAAGTGTGATATGTTCTCTGTTCACTGGCCCAATTCGTGCG 120
QY 365 AGAAGATGGCCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGG 424

```

Db	1653	CGCTTCGTCACATGAAGGACACCTTTGGCTACGACGTATGGCCCGAGAGACCGCCAA	171
Qy	963	GAACCTGTCTGCTGA	978
Db	1713	GAACATCTCCGCTAA	1728
RESULT 7			
LOCUS	AX472800	697 bp	DNA linear
DEFINITION	Sequence 15 from Patent EP1213354.		
ACCESSION	AX472800		
VERSION	AX472800.1	GI:22207649	
KEYWORDS	Escherichia coli		
SOURCE	Escherichia coli		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.		
REFERENCE	1		
AUTHORS	Asako, H., Matsumura, K., Shimizu, M., Ito, N. and Wakita, R.		
TITLE	Process for producing optically active 4-halo-3-hydroxybutanoate		
JOURNAL	Patent: EP 1213354-A 15 12-JUN-2002;		
FEATURES	Sumitomo Chemical Company, Limited (JP)		
source	Location/Qualifiers		
	1..697		
	/organism="Escherichia coli"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:562"		
ORIGIN			
	Query Match	60.2%; Score 588.6; DB 6; Length 697;	
	Best Local Similarity	94.4%; Pred. No. 5.1e-116;	
	Matches	611; Conservative 0; Mismatches 35; Indels 1; Gaps 1;	
Qy	4	TCTAACGGAAGACTTTTACATTCAGCAACGGGTCAAGATTCTCGCGTCGGCTTTGGT	63
Db	52	TCCAAACGGAANACTTTTACACTGAGCAACGGGTCAAAATTCCTGCGTCGGCTTTGGT	111
Qy	64	ACCTTCGCTAGTGAAGTTTCCAAAGGCGAGACCTATCTCTGTCAACCTGCGCTTGAAG	123
Db	112	ACCTNCGCTAGTGAAGTTTCCAAAGGCGAAACCTATNCTCTGTCAACCTGCGCTGAA	171
Qy	124	ACCGTTACCGTCACTTGGACTGTCTGTGTACTACTCTGAAACGAGGTTGAGTTGTGAG	183
Db	172	ACCGTTTACCGTCTNCTTGGACTGTCTGTGTACTACTCTGAAACGAGGTTGAGTTGTGAG	231
Qy	184	GGTATCCGTGACTTCTTCCGAAGGAAACCCCTCGGTGAAGCGTCAAGACATCTTCGCTGC	243
Db	232	GGTNTCCGTGACTTCTTGAAGGAAACCCCTCGGTGAAGCGTCAAGACATCTTCGCTGC	291
Qy	244	ACCAAGTGTGGAAACCACTCCACCGTTATGAGAGAGTCTCTGGTCCATTTGACGACTCC	303
Db	292	ACCAAGTGTGGAAACCACTCCACCGTTATGAGAGAGTCTCTGGTCCATTTGACGACTCC	351
Qy	304	CTGAAGCGTCTTGGACTTGTACTAGTTGATGTCTCTCGTTCACCTGGCCCATTTGCTGC	363
Db	352	CTGAAGCGTCTTGGACTTGTACTAGTTGATGTCTCTCGTTCACCTGGCCCATTTGCTGC	411
Qy	364	GAGAGAAATGGCCAGGTTGAGCCCAAGATTGGCCCTGACGCGCAAAATCGTCTTCTCAAG	423
Db	412	GAAAAAATGGCCAGGTTGAGCCCAAAATTTGGCCCTGACGCGCAAAATCTTCTCAAG	471
Qy	424	GACCTGACGAGAACCCGAGCCACATGCGCGGCTATGGAGAGATTTTATGAGATCCG	483
Db	472	GACCTGACGGA-AAATCCNANCCACCTGGCGGCTATGGAAAAAATTTTNGATGCC	530
Qy	484	AAGGCCAGTCTCATTTGGTGTCTCCAACTGGACCATTTGCCACCTTCAGAGATCCG	543
Db	531	AAGGCCAGTCTCATTTGGTGTCTCCAACTGGACCATTTGCCACCTTCAGAGATCCG	590
Qy	544	TTGCGCAAGGTCATGCTCTCAGCCCAACAGATCGAGATTCACCCCTCTCTGCCAACGAG	603
Db	591	TTNGCGAAGGTCATGCTCTCAGCCCAACAGATCGAGATTCACCCCTCTCTGCCAACGAG	650

Higashihiroshima, Hiroshima 739-0046, Japan
(E-mail: skao_t@rib.go.jp, Tel: 81-824-20-0824, Fax: 81-824-20-0808)

FEATURES

source
Location/Qualifiers
1..1183
/organism="Aspergillus oryzae"
/mol_type="mRNA"
/strain="RIB40"
/db_xref="taxon:5062"
/clone="JK1125"
53..11030
/codon_start=1
/product="reductase-like protein"
/protein_id="BAC56099.1"
/db_xref="GI:28144501"
/translation="MSGKFTLSNGVKIPGVGFTFASGSGKGYQAVTKALERYG
RHLDCAFYLNEDVGGIDFLKKNPSVKREDIFVCTKVNNHLHRPEDEVMSVNSL
KRLDLYVDLFLVHPVASEKEQKPKIGDPGKYVILKELTENPPTWRAMEKIYD
GKAAGVSNWITPGLKPKFAEIKPHVQNIIEHPFLNNELVQFCFNDILPEAYS
PLGSONQVPTTGERVSENKTLNEIAQKGNLTAQVLIAGLRRGVVLPKSNPARI
SNFKSIELTDDEFEAVNKVAEGRHFRFVNKMDTFGYDVWPEETAKNLSA"

ORIGIN

Query Match 75.1%; Score 734.8; DB 8; Length 1183;
Best Local Similarity 84.5%; Pred. No. 1.9e-147;
Matches 826; Conservative 0; Mismatches 152; Indels 0; Gaps 0;
QY 1 ATGTCTAACGGAAGACTTTTACATGAGCAACGGCGTCAAGATTCTCGTGGCTGGCTTT 60
DB 53 ATGTCTAACGGAAGACTTTTACATGAGCAACGGCGTCAAGATTCTCGTGGCTGGCTTT 112
QY 61 GGTACTTTCGTAGTGAAGTTTCAAGGGCGAGACTTACTGTGTGTCACCACTGGCTG 120
DB 113 GGTACTTTCGTAGTGAAGTTTCAAGGGCGAGACTTACTGTGTGTCACCACTGGCTG 172
QY 121 AAGACCGGTTCACGTCTACTGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
DB 173 GAGACCGGATACCGTCACTTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 232
QY 181 GAGGTATTCGTGATCTTCTGAAAGAGAACCCCTCGGTGAGCGTGAAGCACTTCGTC 240
DB 233 GATGATATCCATGACTTCTCAAGAAAGAACCCCTCGGTGAGCGTGAAGCACTTCGTC 292
QY 241 TGCACCAAGGTGTGAACCACTCCACGCTTATGAGGAGCTCTCTGCTGCTGCTGCTGCTG 300
DB 293 TGCACCAAGGTGTGAACCACTCCACGCTTATGAGGAGCTCTCTGCTGCTGCTGCTGCTG 352
QY 301 TCCCTGAAGCGTCTTGGACTTGAATGATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
DB 353 TCCCTGAAGCGTCTTGGACTTGAATGATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 412
QY 361 GCGGAGAGATGCGCAGGTGAGCCCAAGATTGCGCTGAGCGCAATACGTCACTTCTC 420
DB 413 TCCGAG 472
QY 421 AAGACCTTGACGAGAACCCCGAGCCACATGCGCGCTATGAGAGAGATTATGAGGAT 480
DB 473 AAGAGCTCACTGAGAACCCCGAGCCACATGCGCGCTATGAGAGAGATTATGAGGAT 532
QY 481 CGAAGGCGAGGTTCATTTGTTCTCMACTGAGACCATGCGGACCTTGAGAGATGCTC 540
DB 533 GCGAAGGCGAGGTTCATTTGTTCTCMACTGAGACCATGCGGACCTTGAGAGATGCTC 592
QY 541 AAGTTTCGCAAGGTTCATCTCCAGCCCAACAGATTCAGAGATTACCCCTTCTGCTGCTG 600
DB 593 AAGTTTCGCGAGATCAAGCTCATGTCACAGATTGAGATTCACCCCTTCTGCTGCTG 652
QY 601 GAGAGCTGGTGCAGTACTGTTCTTCCAGAACATTAATGCGCGTGGCTTACTTCTCTG 660
DB 653 AACGAGTTGGTCCAGTTCTGTTCTTCAAGAACGACATCTCTCCCGGAGGCTTACTTCG 712
QY 661 GGCTCGAGAACCGAGTTTCCACACCGGTGAGCGGTGTCAGGAGAACAGACTCTGAC 720
DB 713 GGCTCGAGAACCGAGTTTCCACACCGGTGAGCGGTGTCAGGAGAACAGACTCTGAC 772

QY 721 GAGATCGCGAGAGCGGCGCAACACCTTGTCTAGGTTCTTATTGCTGGGCTGCGC 780
DB 773 GAGATCGCGAGAGCGGCGCAACACCTTGTCTAGGTTCTTATTGCTGGGCTGCGC 832
QY 781 CTGCTAGCTGCTGCTTCTCCCAAGAGCTCCCAACCCCAAGCGGATGAGTCCAACTCAAG 840
DB 833 CGCGCTACGTGCTGCTTCTCCCAAGAGCTCCCAACCCCAAGCGGATGAGTCCAACTCAAG 892
QY 841 AGCATTTAGCTCTCGATGCGGACCTTTGAAGCCATCAATGCGCTTGCACCAAGGCTGCTCAC 900
DB 893 AGCATTTAGCTCTCGATGCGGACCTTTGAAGCCATCAATGCGCTTGCACCAAGGCTGCTCAC 952
QY 901 TTCCGCTTCTGCTCAACATGAAGGATCTTTTGGATATGATGCTGCGCCCGAGAGACCGCC 960
DB 953 TTCCGCTTCTGCTCAACATGAAGGATCTTTTGGATATGATGCTGCGCCCGAGAGACCGCC 1012
QY 961 AAGAACCTGCTGCTGCTGA 978
DB 1013 AAGAACCTGCTGCTGCTGA 1030

RESULT 6
ANI496025 2397 bp DNA linear PLN 17-JUL-2003
LOCUS Aspergillus nidulans gldB gene for NADP(+) -dependent glycerol
DEFINITION dehydrogenase, exons 1-2.
ACCESSION AJ496025
VERSION AJ496025.1 GI:32879319
KEYWORDS gldB gene; NADP(+) -dependent glycerol dehydrogenase.
SOURCE Emericella nidulans (anamorph: Aspergillus nidulans)
ORGANISM Emericella nidulans
REFERENCE 1
AUTHORS de Vries, R.P.
TITLE Glycerol dehydrogenase, encoded by gldB is essential for
osmotolerance in Aspergillus nidulans
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2397)
AUTHORS de Vries, R.P.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2002) de Vries R.P., Microbiology, Utrecht
University, Padualaan 8, Utrecht, 3584 CH, NETHERLANDS
FEATURES
Source
1..2397
/organism="Emericella nidulans"
/mol_type="genomic DNA"
/db_xref="taxon:162425"
693..1728
/gene="gldB"
join(693..1100,1159..1728)
/gene="gldB"
/EC_number="1.1.1.72"
/function="glycerol biosynthesis"
/codon_start=1
/product="NADP(+) -dependent glycerol dehydrogenase"
/protein_id="CAD42649.1"
/db_xref="GI:32879320"
/translation="MSGKFTLSNGVTIPAVGFTFASGSGKGYQAVTKALEVGY
RHLDCAFYLNEDVGGIDFLKKNPSVKREDIFVCTKVNNHLHRPEDEVMSVNSL
KRLDLYVDLFLVHPVASEKEQKPKIGDPGKYVILKELTENPPTWRAMEKIYD
GKAAGVSNWITPGLKPKFAEIKPHVQNIIEHPFLNNELVQFCFNDILPEAYS
PLGSONQVPTTGERVSENKTLNEIAQKGNLTAQVLIAGLRRGVVLPKSNPARI
SNFKSIELTDDEFEAVNKVAEGRHFRFVNKMDTFGYDVWPEETAKNLSA"
-693..1100
/gene="gldB"
/number=1
1101..1158
/gene="gldB"
/number=1
1159..1728
/gene="gldB"
exon
intron
exon

Db 181 GAGGGTATCCGTGACTCTCTGAAGGAAACCCCTCGGTGAAGCGTGAGGACATCTTCGTG 240
Qy 241 TGCAACAGGTGTGGAAACACCTCCACCGTTATGAGAGCGTCTCTGGTCCATTGACGAC 300
Db 241 TGCACCAAGGTGTGGAACACCTCCACCGTTATGAGGAGCGTCTCTGGTCCATTGACGAC 300
Qy 301 TCCCTGAAGCGTCTTGACCTGACTAGCTAGCTGTGATATGTTCTCTGTTCACTGGCCCATGCT 360
Db 301 TCCCTGAAGCGTCTTGACCTGACTAGCTAGCTGTGATATGTTCTCTGTTCACTGGCCCATGCT 360
Qy 361 GCCGAGAGAAATGCCAGGAGTGAAGCCCAAGTATGGCCCTCAGCGCAAAATACGTCATTCTC 420
Db 361 GCCGAGAGAAATGCCAGGAGTGAAGCCCAAGTATGGCCCTCAGCGCAAAATACGTCATTCTC 420
Qy 421 AAGGACCTGACCGAGAAACCCCGAGCCCAATGCGCGCTATGAGAGAAATTTATGAGGAT 480
Db 421 AAGGACCTGACCGAGAAACCCCGAGCCCAATGCGCGCTATGAGAGAAATTTATGAGGAT 480
Qy 481 CGCAAGGCGAGTCCATTGCTGCTCCAACTGACCAATTCGCCACCTTGAGAGAGATGTC 540
Db 481 CGCAAGGCGAGTCCATTGCTGCTCCAACTGACCAATTCGCCACCTTGAGAGAGATGTC 540
Qy 541 AAGTTCCGCAAGGTCATGCTCAGCCCAACAGATCGAGATTCAACCCCTTCTGCCCCAAC 600
Db 541 AAGTTCCGCAAGGTCATGCTCAGCCCAACAGATCGAGATTCAACCCCTTCTGCCCCAAC 600
Qy 601 GAGGAGCTGGTGCAGTACTGCTTCTCAAGAACATTTATGCGCGTACTCTCTCTG 660
Db 601 GAGGAGCTGGTGCAGTACTGCTTCTCAAGAACATTTATGCGCGTACTCTCTCTG 660
Qy 661 GGCTCGAGAACAGGTTCCACACCGGTGAGCGGTGAGCGAGAACAGATCTCTGAAC 720
Db 661 GGCTCGAGAACAGGTTCCACACCGGTGAGCGGTGAGCGAGAACAGATCTCTGAAC 720
Qy 721 GAGATCCCGAGAGGCGGCAACACCTTGTCTGAGGTTCTTATGCTGGGTCTGCGC 780
Db 721 GAGATCCCGAGAGGCGGCAACACCTTGTCTGAGGTTCTTATGCTGGGTCTGCGC 780
Qy 781 CGTGGCTACGTCTCTCCAGAGCTCCAAACCCAGCGCATGAGTCCAACTTCAAG 840
Db 781 CGTGGCTACGTCTCTCCAGAGCTCCAAACCCAGCGCATGAGTCCAACTTCAAG 840
Qy 841 AGCATGAGCTCTCCGATGCGGACTTTGAAGCCATCAATCCGTTGCCAAGGCTGCTAC 900
Db 841 AGCATGAGCTCTCCGATGCGGACTTTGAAGCCATCAATCCGTTGCCAAGGCTGCTAC 900
Qy 901 TTCGTTTCGTCACATGAGGATCTTTCGGATATGATCTGGCCGAGGAGCGCC 960
Db 901 TTCGTTTCGTCACATGAGGATCTTTCGGATATGATCTGGCCGAGGAGCGCC 960
Qy 961 AAGAACCTGCTGCGTGA 978
Db 961 AAGAACCTGCTGCGTGA 978

RESULT 3

AX815044
LOCUS 978 bp DNA linear PAT 05-DEC-2003
DEFINITION Sequence 4 from Patent EP1323827.
ACCESSION AX815044
VERSION AX815044.1 GI:39104167
KEYWORDS Penicillium citrinum
SOURCE Penicillium citrinum
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
REFERENCE 1
AUTHORS Asako,H., Wakita,R. and Itoh,N.
TITLE Method for producing optically active
JOURNAL 2-hydroxycycloalkane carboxylic acid ester
Patent: EP 1323827-A 4 02-JUL-2003;
Sumitomo Chemical Company, Limited (JP)
FEATURES Location/Qualifiers

source 1..978
/organism="Penicillium citrinum"
/mol_type="unassigned DNA"
/db_xref="taxon:5077"
CDS 1..978
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB83559.1"
/translation="MSNGKFTLSNGKIPGVGFGTASBGSKGETVTATVTTALKTGY
RHLCAWYLNVEGEVIGIRDFLENPSEVRKEDIFVCTKWNHILHVEDVLSIDSL
KRLGIDYVDFMELVHPWIAAEKNGGEPKIGDKYVILKDLTENPEPTWRAMEKIYD
RKASIGVSNVTIADLEKMSKFAKMPHANOIEIHPFLPNEBELVOYCFKSNIMFVAYS
PLGSONQVPTTGERSVSKNTLINEAEKGNLTAVLIANGLRGVVLPKSNPKRIE
SNFKSIELSDADFEAINNAVKGRHFRFVNMDTFTGYDWPETAKNLSA"

ORIGIN

Query Match 100.0%; Score 978; DB 6; Length 978;
Best Local Similarity 100.0%; Pred. No. 1.1e-199;
Matches 978; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGTCTAACGGAAGACTTTTACATTGAGCAACGGCGTCAAGATTCTTGGCGTGGCTTT 60
Db 1 ATGTCTAACGGAAGACTTTTACATTGAGCAACGGCGTCAAGATTCTTGGCGTGGCTTT 60
Qy 61 GGTACTCTTCGTAGTGAAGGTTTCAAGGGCGAGACCTATCTGCTGTCACCACTGCCCTG 120
Db 61 GGTACTCTTCGTAGTGAAGGTTTCAAGGGCGAGACCTATCTGCTGTCACCACTGCCCTG 120
Qy 121 AAGACCGGTTACCGTCACTTGGACTGTGCTTGGTATCTACCTGAACGAGGGTGAGTTGGT 180
Db 121 AAGACCGGTTACCGTCACTTGGACTGTGCTTGGTATCTACCTGAACGAGGGTGAGTTGGT 180
Qy 181 GAGGTTATCCGTGACTTCTTGAAGGAGAACCCCTCGGTGAAGCGTGAGGACATCTTCGTC 240
Db 181 GAGGTTATCCGTGACTTCTTGAAGGAGAACCCCTCGGTGAAGCGTGAGGACATCTTCGTC 240
Qy 241 TGCACCAAGGTGTGGAACCACTCCACCGTTATGAGGAGCGTCTTGGTCCATTGACGAC 300
Db 241 TGCACCAAGGTGTGGAACCACTCCACCGTTATGAGGAGCGTCTTGGTCCATTGACGAC 300
Qy 301 TCCCTGAAGCGTCTTGGACTTGAATGATATGTTCTCTGTTCACTGGCCCATGCT 360
Db 301 TCCCTGAAGCGTCTTGGACTTGAATGATATGTTCTCTGTTCACTGGCCCATGCT 360
Qy 361 GCCGAGAGAAATGCGCAGGCGTGAAGGATGGCCCTGACGGCAAAATACGTCATTCTC 420
Db 361 GCCGAGAGAAATGCGCAGGCGTGAAGGATGGCCCTGACGGCAAAATACGTCATTCTC 420
Qy 421 AAGGACCTGACCGAGAAACCCGAGCCCAATGCGCGCTATGAGAGAAATTTATGAGGAT 480
Db 421 AAGGACCTGACCGAGAAACCCGAGCCCAATGCGCGCTATGAGAGAAATTTATGAGGAT 480
Qy 481 CGCAAGGCGAGTCCATTGCTGCTCCAACTGAGCATTGGCCGAGGAGATGTC 540
Db 481 CGCAAGGCGAGTCCATTGCTGCTCCAACTGAGCATTGGCCGAGGAGATGTC 540
Qy 541 AAGTTCCGCAAGGTCATGCTTCTCAAGAACATTTATGCGCGTGGCTACTCTCTCTG 600
Db 541 AAGTTCCGCAAGGTCATGCTTCTCAAGAACATTTATGCGCGTGGCTACTCTCTCTG 600
Qy 601 GAGGAGCTGGTGCAGTACTGCTTCTCAAGAACATTTATGCGCGTGGCTACTCTCTCTG 660
Db 601 GAGGAGCTGGTGCAGTACTGCTTCTCAAGAACATTTATGCGCGTGGCTACTCTCTCTG 660
Qy 661 GGCTCGAGAACAGGTTCCACACCGGTGAGCGGTGAGCGAGAACAGATCTCTGAAC 720
Db 661 GGCTCGAGAACAGGTTCCACACCGGTGAGCGGTGAGCGAGAACAGATCTCTGAAC 720
Qy 721 GAGATCCCGAGAGGCGGCAACACCTTGTCTGAGGTTCTTATGCTGGGTCTGCGC 780
Db 721 GAGATCCCGAGAGGCGGCAACACCTTGTCTGAGGTTCTTATGCTGGGTCTGCGC 780

Sumitomo Chemical Company, Limited (JP)
Location/Qualifiers
1. .978
/organism="Penicillium citrinum"
/mol_type="unassigned DNA"
/db_xref="taxon:5077"
1. .978
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD43580.1"
/db_xref="GI:22207661"
/db_xref="REMTREMBL:CAD43580"
/translation="MSGKFTLSNGVKIPGVGFTFASGSGKETYVAVTTALKTGY
RHLDCAWYLNEGEVGIKDFLENPSVKREDIFVCTKVNHLHRYEDVLSIDSL
KRLGLDYVDFLWHPFAAEKNGQGEPIKIPDGKYLKDLTENPFTWRAMEKIYED
RKARSIGVSNWTADLEKMSKFAKVMPHANOIEIHPFLNEBELVOYCFSKNIMPVAYS
PLGSONOVPTTGERVSENKTLNEIAEKGNLTALQVLIAMGLRRGVVLKPSNPKRIE
SNFKSIELSDADFEAINAVAKGRHFRFVNMKDTFGYDWPETAKLSA"

FEATURES
Source
CDS
ORIGIN

Query Match 100.0%; Score 978; DB 6; Length 978;
Best Local Similarity 100.0%; Pred. No. 1.1e-199;
Matches 978; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTAACGGAAGACTTTCACATTGAGCAACGGCGTCAAGATTCTCGCGTGGCTTT 60
Db 1 ATGCTAACGGAAGACTTTCACATTGAGCAACGGCGTCAAGATTCTCGCGTGGCTTT 60

QY 61 GGTACCTTCGCTAGTGAAGTTTCAAGGGCGAGACCTATCTGCTGTACCACTGCCCTG 120
Db 61 GGTACCTTCGCTAGTGAAGTTTCAAGGGCGAGACCTATCTGCTGTACCACTGCCCTG 120

QY 121 AAGACCGGTTACCGTCACTTGCCTGGTACTACTGAAAGGGTGAAGGTTGAGTTGGT 180
Db 121 AAGACCGGTTACCGTCACTTGCCTGGTACTACTGAAAGGGTGAAGGTTGAGTTGGT 180

QY 181 GAGGCTATCCGTGACTTCTGAAAGGAGAACCCCTCGGTGAAGCGTGAAGACATCTTCGTC 240
Db 181 GAGGCTATCCGTGACTTCTGAAAGGAGAACCCCTCGGTGAAGCGTGAAGACATCTTCGTC 240

QY 241 TGCACCAAGGTGTGAACCACTCCACGGTTATGAGGACGTCCTCTGTCCTCATTTGACG 300
Db 241 TGCACCAAGGTGTGAACCACTCCACGGTTATGAGGACGTCCTCTGTCCTCATTTGACG 300

QY 301 TCCCTGAAGCGCTTGGACTTACTGCTTATGATTTCTGTTCTGTTCACTGGCCCATTTGCT 360
Db 301 TCCCTGAAGCGCTTGGACTTACTGCTTATGATTTCTGTTCTGTTCACTGGCCCATTTGCT 360

QY 361 CCCGAGAAGATAGCGGAGTGAAGGATTCGCGCTGACGGCAAAATACGTCATTCTC 420
Db 361 CCCGAGAAGATAGCGGAGTGAAGGATTCGCGCTGACGGCAAAATACGTCATTCTC 420

QY 421 AAGGACCTGACGAGAACCCCGAGCCACATGGCGCGCTATGGAGAGATTTATGAGAT 480
Db 421 AAGGACCTGACGAGAACCCCGAGCCACATGGCGCGCTATGGAGAGATTTATGAGAT 480

QY 481 CGCAAGCCAGGTCATTGTTGTCTCCAACTGACCACTTGCCTGCGGCTTGAAGATGTC 540
Db 481 CGCAAGCCAGGTCATTGTTGTCTCCAACTGACCACTTGCCTGCGGCTTGAAGATGTC 540

QY 541 AAGTTCGCAAGGTATGCTTCAGCCCAACAGATCGAGATTACCCCTTCCTGCCCAAC 600
Db 541 AAGTTCGCAAGGTATGCTTCAGCCCAACAGATCGAGATTACCCCTTCCTGCCCAAC 600

QY 601 GAGGAGCTGGTGCAGTACTGTTCTTCCAAAGAACATTATGCCCGTGGCTACTCTCTCTG 660
Db 601 GAGGAGCTGGTGCAGTACTGTTCTTCCAAAGAACATTATGCCCGTGGCTACTCTCTCTG 660

QY 661 GGCTCGAGAACAGGTTCCACACACCGGTGAGCGGGTACGCGAACAAGATCTCTGAAC 720
Db 661 GGCTCGAGAACAGGTTCCACACACCGGTGAGCGGGTACGCGAACAAGATCTCTGAAC 720

QY 721 GAGATCGCCGAGAAGGGGGGCAACACCTTCTCAGGTTCTTATTGCTCGGGTCTCGCG 780
Db 721 GAGATCGCCGAGAAGGGGGGCAACACCTTCTCAGGTTCTTATTGCTCGGGTCTCGCG 780

Db 721 GAGATCGCCGAGAAGGGGGGCAACACCTTCTCAGGTTCTTATTGCTCGGGTCTCGCG 780
QY 781 CGTGGCTAGCTGTTCTCCCAAGAGCTTCAACCCCAAGCGCATTTAGTCCAATCTCAAG 840
Db 781 CGTGGCTAGCTGTTCTCCCAAGAGCTTCAACCCCAAGCGCATTTAGTCCAATCTCAAG 840
QY 841 AGCATTTAGCTCTCCGATGCCGACTTTGAAGCCATCAATGCCGTTGCCAAGGGTCTGCAC 900
Db 841 AGCATTTAGCTCTCCGATGCCGACTTTGAAGCCATCAATGCCGTTGCCAAGGGTCTGCAC 900
QY 901 TTCCGTTTCTCAACATGAAGATACCTTCGATATGATGTCGCCCGAGAGACCCGCC 960
Db 901 TTCCGTTTCTCAACATGAAGATACCTTCGATATGATGTCGCCCGAGAGACCCGCC 960
QY 961 AAGAACCCTGCTGCGGTGA 978
Db 961 AAGAACCCTGCTGCGGTGA 978

RESULT 2
AX472815
LOCUS
DEFINITION
AX472815
ACCESSION
AX472815.1 GI:22207666
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Source
CDS

AX472815
Sequence 30 from Patent EPI213354.
AX472815.1 GI:22207666
Penicillium citrinum
Penicillium citrinum
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
1
Asako, H., Matsumura, K., Shimizu, M., Ito, N. and Wakita, R.
Process for producing optically active 4-halo-3-hydroxybutanoate
Patent: EP 1213354-A 30 12-JUN-2002;
Sumitomo Chemical Company, Limited (JP)
Location/Qualifiers
1. .978
/organism="Penicillium citrinum"
/mol_type="unassigned DNA"
/db_xref="taxon:5077"
1. .978
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD43583.1"
/db_xref="GI:22207667"
/db_xref="REMTREMBL:CAD43583"
/translation="MSGKFTLSNGVKIPGVGFTFASGSGKETYVAVTTALKTGY
RHLDCAWYLNEGEVGIKDFLENPSVKREDIFVCTKVNHLHRYEDVLSIDSL
KRLGLDYVDFLWHPFAAEKNGQGEPIKIPDGKYLKDLTENPFTWRAMEKIYED
RKARSIGVSNWTADLEKMSKFAKVMPHANOIEIHPFLNEBELVOYCFSKNIMPVAYS
PLGSONOVPTTGERVSENKTLNEIAEKGNLTALQVLIAMGLRRGVVLKPSNPKRIE
SNFKSIELSDADFEAINAVAKGRHFRFVNMKDTFGYDWPETAKLSA"

Query Match 100.0%; Score 978; DB 6; Length 978;
Best Local Similarity 100.0%; Pred. No. 1.1e-199;
Matches 978; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTAACGGAAGACTTTCACATTGAGCAACGGCGTCAAGATTCTCGCGTGGCTTT 60
Db 1 ATGCTAACGGAAGACTTTCACATTGAGCAACGGCGTCAAGATTCTCGCGTGGCTTT 60

QY 61 GGTACCTTCGCTAGTGAAGTTTCAAGGGCGAGACCTATCTGCTGTACCACTGCCCTG 120
Db 61 GGTACCTTCGCTAGTGAAGTTTCAAGGGCGAGACCTATCTGCTGTACCACTGCCCTG 120

QY 121 AAGACCGGTTACCGTCACTTGCCTGGTACTACTGAAAGGGTGAAGGTTGAGTTGGT 180
Db 121 AAGACCGGTTACCGTCACTTGCCTGGTACTACTGAAAGGGTGAAGGTTGAGTTGGT 180

QY 181 GAGGCTATCCGTGACTTCTGAAAGGAGAACCCCTCGGTGAAGCGTGAAGACATCTTCGTC 240
Db 181 GAGGCTATCCGTGACTTCTGAAAGGAGAACCCCTCGGTGAAGCGTGAAGACATCTTCGTC 240

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2004, 11:12:09 ; Search time 3874 Seconds
(without alignments)
10942.046 Million cell updates/sec

Title: US-10-004-115B-2

Perfect score: 978
Sequence: 1 atgtctaaccgaagaagacttt.....ccaagaacctgtctgcgtga 978

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.on.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.mu.*

21: em.ov.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	978	100.0	978	6	AX472787	AX472787 Sequence
2	978	100.0	978	6	AX472815	AX472815 Sequence
3	978	100.0	978	6	AX815044	AX815044 Sequence
4	978	100.0	996	6	AX472813	AX472813 Sequence
5	734.8	75.1	1183	8	AB079030	AB079030 Aspergill
6	607.6	62.1	2397	8	ANI495025	ANI495025 Aspergill
7	588.6	60.2	597	6	AX472800	AX472800 Sequence
8	471	48.2	762	8	AB014493	AB014493 Gibberell
9	417	42.6	417	6	AX472807	AX472807 Sequence
10	412	42.1	743	6	AX472804	AX472804 Sequence
11	265	27.1	331	6	AX472803	AX472803 Sequence
12	247	25.3	720	8	CNS01C4M	AL112097 Botrytis
13	233.2	23.8	720	8	CNS01SUI	AL112097 Botrytis
14	230.6	23.6	660	8	CNS01BYU	AL114862 Botrytis
15	226.2	23.1	780	8	CNS01BQ2	AL114846 Botrytis
16	208	21.3	720	8	CNS01AFA	AL112862 Botrytis
17	199.6	20.4	1055	6	E11242	E11242 cdna encodi
18	188.8	19.3	720	8	CNS01DNY	AL117062 Botrytis
19	180.2	18.4	939	8	AY033888	AY033888 Magnapor
20	178.8	18.3	1123	3	BT003280	BT003280 Drosophi
21	174.8	17.9	780	8	CNS01DM8	AL117000 Botrytis
22	144.2	14.7	1121	6	AX470403	AX470403 Sequence
23	144.2	14.7	1121	6	BD177581	BD177581 Erythro
24	144.2	14.7	1121	6	BD177586	BD177586 Erythro
25	142.6	14.6	1077	6	AX470402	AX470402 Sequence
26	142.6	14.6	1077	6	BD177580	BD177580 Erythro
27	142.6	14.6	1077	6	BD177585	BD177585 Erythro
28	141.6	14.5	456	8	CNS01B1R	AL113671 Botrytis
29	137.4	14.0	2359	4	BOVALRDB	M59754 Bovine lens
30	126	12.9	970	4	FIGALDRD	L14950 Sus scrofa
31	126	12.9	970	4	SSU46065	U46065 Sus scrofa
32	124.8	12.8	1156	4	BOVALDRD	M31463 Bovine aldo
33	124.8	12.8	1187	4	S54973	S54973 20 alpha-hy
34	122.2	12.5	456	8	CNS01BBK	AL114024 Botrytis
35	121.2	12.4	2309	8	SSU26463	U26463 Sporidobol
36	121	12.4	1395	10	BC051128	BC051128 Mus muscu
37	119.8	12.2	1310	4	OCU12316	U12316 Oryctolagus
38	119.8	12.2	1494	4	RABALRE	M32818 Rabbit (clo
39	119.2	12.2	1335	6	AR380465	AR380465 Sequence
40	119.2	12.2	1335	9	HUMARXA	M34720 Human aldos
41	119.2	12.2	1354	9	HUMALRMB	J05474 Human aldos
42	119.2	12.2	1355	9	AF328729	AF328729 Homo sapi
43	119.2	12.2	1367	6	AX440479	AX440479 Sequence
44	119.2	12.2	1367	9	HSALRE	X15414 Human mRNA
45	119.2	12.2	1368	6	AX451813	AX451813 Sequence

ALIGNMENTS

RESULT 1	AX472787	978 bp	DNA	linear	PAT 09-AUG-2002
LOCUS	AX472787	Sequence 2 from Patent EP1213354.			
DEFINITION	AX472787				
ACCESSION	AX472787.1	GI:22207640			
VERSION	AX472787.1				
KEYWORDS	Penicillium citrinum				
SOURCE	Penicillium citrinum				
ORGANISM	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;				
	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.				
REFERENCE	1				
AUTHORS	Asako, H., Matsumura, K., Shimizu, M., Ito, N. and Wakita, R.				
TITLE	Process for producing optically active 4-halo-3-hydroxybutanoate				
JOURNAL	Patent: EP 1213354-A 2 12-JUN-2002;				

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER
 ; FILE REFERENCE: 210121.478C14
 ; CURRENT APPLICATION NUMBER: US/09/702,705
 ; CURRENT FILING DATE: 2000-10-30
 ; NUMBER OF SEQ ID NOS: 1833
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 340
 ; LENGTH: 316
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-702-705-340

Query Match 28.1%; Score 487; DB 4; Length 316;
 Best Local Similarity 36.6%; Pred. No. 8.8e-44;
 Matches 118; Conservative 58; Mismatches 122; Indels 24; Gaps 8;

QY	6	TFT-LSNGVKIPGVGFGTFASEGSKGETYTAVTTALKTGVRHLCDAWYVNLNVEGVGEIR	64
DB	3	TFVELSTKAKMPVIGLGTWKS--PLGKVEAVKVAIDAGYRHIDCAVYVQNEHEVGEAIQ	60
QY	65	DFLKENPSVKREDIFVCTKVNHLHRYEDVLWSIDSLKRLGLDYVDMFLVHWPAAEKN	124
DB	61	EKIQEK-AVKREDLFIVSKLWPTFFERPLVRKAPEKTLKDLKSLVDVLIHWPQGFSG	119
QY	125	GQGEPK-----IGPDGKVIKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEK	178
DB	120	DDLFPKDDKGNAG--GKATFL-----DAWEAMELVDEGLVKALGVSNFHFQIEK	169
QY	179	M--SKFAKVMPHANQIEHPFLPNEELVOYCFSKNIMPVAYSPLGSONQVPTTGERVS--	234
DB	170	LNKPKGLKYKPVTVNQVECHPYLTQEKLIQYCHSKGITVTAYSPGSPDRPWAKPDPSLL	229
QY	235	ENKTLNEIAEKGNTLAQVLIAGLRRGYVWLPKSSNPKRIESNFKSIELSDADFEAINA	294
DB	230	EDPKIEIAAKHKHTAAQVLIHFHQRNVIVIPKSVTPARIVENTQVDFKLSDEEMATI	289
QY	295	VAKGRHFRFVNKMDTFGYDWP	316
DB	290	LSFNRNWRACNVLSQSHLEDYP	311

RESULT 14
 US-09-736-457-340
 ; Sequence 340, Application US/09736457
 ; Patent No. 6509448
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Lodes, Michael A.
 ; APPLICANT: Fanger, Gary
 ; APPLICANT: Vedwick, Tom
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Retter, Marc
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Fan, Liqun
 ; APPLICANT: Wang, Aijun
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER
 ; FILE REFERENCE: 210121.478C15
 ; CURRENT APPLICATION NUMBER: US/09/736,457
 ; CURRENT FILING DATE: 2000-12-13
 ; NUMBER OF SEQ ID NOS: 1864
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 340
 ; LENGTH: 316
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-736-457-340

Query Match 28.1%; Score 487; DB 4; Length 316;
 Best Local Similarity 36.6%; Pred. No. 8.8e-44;
 Matches 118; Conservative 58; Mismatches 122; Indels 24; Gaps 8;

QY 6 TFT-LSNGVKIPGVGFGTFASEGSKGETYTAVTTALKTGVRHLCDAWYVNLNVEGVGEIR 64
 DB 3 TFVELSTKAKMPVIGLGTWKS--PLGKVEAVKVAIDAGYRHIDCAVYVQNEHEVGEAIQ 60
 QY 65 DFLKENPSVKREDIFVCTKVNHLHRYEDVLWSIDSLKRLGLDYVDMFLVHWPAAEKN 124
 DB 61 EKIQEK-AVKREDLFIVSKLWPTFFERPLVRKAPEKTLKDLKSLVDVLIHWPQGFSG 119
 QY 125 GQGEPK-----IGPDGKVIKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEK 178
 DB 120 DDLFPKDDKGNAG--GKATFL-----DAWEAMELVDEGLVKALGVSNFHFQIEK 169
 QY 179 M--SKFAKVMPHANQIEHPFLPNEELVOYCFSKNIMPVAYSPLGSONQVPTTGERVS-- 234
 DB 170 LNKPKGLKYKPVTVNQVECHPYLTQEKLIQYCHSKGITVTAYSPGSPDRPWAKPDPSLL 229
 QY 235 ENKTLNEIAEKGNTLAQVLIAGLRRGYVWLPKSSNPKRIESNFKSIELSDADFEAINA 294
 DB 230 EDPKIEIAAKHKHTAAQVLIHFHQRNVIVIPKSVTPARIVENTQVDFKLSDEEMATI 289
 QY 295 VAKGRHFRFVNKMDTFGYDWP 316
 DB 290 LSFNRNWRACNVLSQSHLEDYP 311

RESULT 15
 US-09-614-124B-340
 ; Sequence 340, Application US/09614124B
 ; Patent No. 6630574
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Lodes, Michael A.
 ; APPLICANT: Fanger, Gary
 ; APPLICANT: Vedwick, Tom
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Retter, Marc
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Mannion, Jane
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
 ; FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER
 ; FILE REFERENCE: 210121.478C9
 ; CURRENT APPLICATION NUMBER: US/09/614,124B
 ; CURRENT FILING DATE: 2001-07-11
 ; NUMBER OF SEQ ID NOS: 1668
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 340
 ; LENGTH: 316
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-614-124B-340

Query Match 28.1%; Score 487; DB 4; Length 316;
 Best Local Similarity 36.6%; Pred. No. 8.8e-44;
 Matches 118; Conservative 58; Mismatches 122; Indels 24; Gaps 8;

QY	6	TFT-LSNGVKIPGVGFGTFASEGSKGETYTAVTTALKTGVRHLCDAWYVNLNVEGVGEIR	64
DB	3	TFVELSTKAKMPVIGLGTWKS--PLGKVEAVKVAIDAGYRHIDCAVYVQNEHEVGEAIQ	60
QY	65	DFLKENPSVKREDIFVCTKVNHLHRYEDVLWSIDSLKRLGLDYVDMFLVHWPAAEKN	124
DB	61	EKIQEK-AVKREDLFIVSKLWPTFFERPLVRKAPEKTLKDLKSLVDVLIHWPQGFSG	119
QY	125	GQGEPK-----IGPDGKVIKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEK	178
DB	120	DDLFPKDDKGNAG--GKATFL-----DAWEAMELVDEGLVKALGVSNFHFQIEK	169
QY	179	M--SKFAKVMPHANQIEHPFLPNEELVOYCFSKNIMPVAYSPLGSONQVPTTGERVS--	234
DB	170	LNKPKGLKYKPVTVNQVECHPYLTQEKLIQYCHSKGITVTAYSPGSPDRPWAKPDPSLL	229
QY	235	ENKTLNEIAEKGNTLAQVLIAGLRRGYVWLPKSSNPKRIESNFKSIELSDADFEAINA	294

Query Match 28.9%; Score 499.5; DB 4; Length 316;
Best Local Similarity 39.4%; Pred. No. 4e-45; Mismatches 106; Indels 29; Gaps 10;
Matches 121; Conservative 51;
Db 9 LNSGVKIPGVGFGTFASEGSGKETVAVTTALKTGYRHLDCAWYILNEGEVGEIGRDLFLK 68
Db 7 LNSGVKIPGVGFGTFASEGSGKETVAVTTALKTGYRHLDCAWYILNEGEVGEIGRDLFLK 64
Qy 69 ENPSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYVDMFLVHWPAAEKNGQGE 128
Db 65 EQ-VVQRDLFLVSKLWCTFDQSMVKGACOKTJSDQLDLYDLIYLIHWPTGF----- 116
Qy 129 PKXIPDGKXVILKDLTENPEP-----TWAMEXIVYEDRKARSIGVSNWTIADLEKXV-- 179
Db 117 -KPGPD--YFPL-DASGNVPSDFTDVTWTAMEQLVDEGLVKAIGVSNFNPQLQIERILN 172
Qy 180 SKFAKMPHANOIEIHPPLNEELVOYCFKSNIMPVAYSPGSONQVPTTGERVS--ENK 237
Db 173 KPLGLKYPANQIECHPVLTOEKLIEYCHKGIVVTAYSPGSDPRPWAKPEDPSLLEDP 232
Qy 238 TLNEIAEKGNTLAQVLIAGLRRGYVVLKSSNPKRIESNFK--SIELSDADFEAINAV 295
Db 233 RIKEIAAKYNTTAQVLIARFPIQRLNVLVIPSVTPIAENFKYVDFELSDNEDMATL--L 290
Qy 296 AKGRHFR 302
Db 291 SYNRRNR 297

RESULT 11
US-09-134-000C-6059
Query Match 28.8%; Score 498; DB 4; Length 298;
Best Local Similarity 38.9%; Pred. No. 5.2e-45; Mismatches 97; Indels 40; Gaps 7;
Matches 118; Conservative 48;
Db 1 MSGKTFTLNSGVKIPGVGFGTFASEGSGKETVAVTTALKTGYRHLDCAWYILNEGEV 60
Db 23 MTLDTKXLANGNEMPRGLGVMRVEDG-SEATNSVKWALEVGYRLIDTAAYVKNVGVG 81
Qy 61 EGIRDFLXNSVREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYVDMFLVHWPAA 120
Db 82 EGIR-----QSGIPREEIFVTTKLNEDQGYENAIKAFDLSLEKGLDYVLYLIHWPTA 136
Qy 121 AEKNGQGEFKIPDGKXVILKDLTENPEPTWRAMEKIVYEDRKARSIGVSNWTIADLEKMS 180
Db 137 -----GKY-----KESWKALEELIYASGRAKAIGVSNFQHHLBDLL 172
Qy 181 KFAKMPHANOIEIHPPLNEELVOYCFKSNIMPVAYSPGSONQVPTTGERVSENKTLN 240
Db 173 TEANVPVMDQIELHTLTQSEPLRKYLAENKIAVEAWSPLGQ-----GNLLSE-PVLV 224
Qy 241 BIAKGGNTLAQVLIAGLRRGYVVLKSSNPKRIESNFK--SIELSDADFEAINAVAKG 298

Query Match 28.8%; Score 498; DB 4; Length 298;
Best Local Similarity 38.9%; Pred. No. 5.2e-45; Mismatches 97; Indels 40; Gaps 7;
Matches 118; Conservative 48;
Db 1 MSGKTFTLNSGVKIPGVGFGTFASEGSGKETVAVTTALKTGYRHLDCAWYILNEGEV 60
Db 23 MTLDTKXLANGNEMPRGLGVMRVEDG-SEATNSVKWALEVGYRLIDTAAYVKNVGVG 81
Qy 61 EGIRDFLXNSVREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYVDMFLVHWPAA 120
Db 82 EGIR-----QSGIPREEIFVTTKLNEDQGYENAIKAFDLSLEKGLDYVLYLIHWPTA 136
Qy 121 AEKNGQGEFKIPDGKXVILKDLTENPEPTWRAMEKIVYEDRKARSIGVSNWTIADLEKMS 180
Db 137 -----GKY-----KESWKALEELIYASGRAKAIGVSNFQHHLBDLL 172
Qy 181 KFAKMPHANOIEIHPPLNEELVOYCFKSNIMPVAYSPGSONQVPTTGERVSENKTLN 240
Db 173 TEANVPVMDQIELHTLTQSEPLRKYLAENKIAVEAWSPLGQ-----GNLLSE-PVLV 224
Qy 241 BIAKGGNTLAQVLIAGLRRGYVVLKSSNPKRIESNFK--SIELSDADFEAINAVAKG 298

Db 225 KIGEKYGAAQVLIIRWHLQNDIIVIPKSVHEKRIQENFAVDFDFELTPEITAINQLYKD 284
Qy 299 RHF 301
Db 285 HRF 287
RESULT 12
US-09-347-803-12
Query Match 28.6%; Score 494.5; DB 3; Length 308;
Best Local Similarity 36.4%; Pred. No. 1.3e-44; Mismatches 114; Conservative 61; Indels 15; Gaps 8;
Matches 114;
Qy 9 LNSGVKIPGVGFGTFASEGSGKETVAVTTALKTGYRHLDCAWYILNEGEVGEIGRDLFLK 68
Db 4 LNSGHTMPSVGLGVWEMDSP--AIRDLHLSALRIGYRHFDCADYQNEAEVGDALAEAF- 60
Qy 69 ENPSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYVDMFLVHWPAAEKNGQGE- 127
Db 61 QTLGVKREDLFTITKLWNSDHGH--VVEACKOSLKLRLDYLDLYLIHPVATRTGTGT 118
Qy 128 -EPKIGPDGKXVILKDLTENPEPTWRAMEKIVYEDRKARSIGVSNWTIADLEKMSKFAKVM 186
Db 119 TASALGDDG--VLDIDTTISLETTHAMEDIVSMGLVRSIGISNYDIFLTRCLAYAKIK 176
Qy 187 PHANQIEIHPPLNEELVOYCFKSNIMPVAYSPGSONQVPTTGERVS--ENKTLNEIAE 244
Db 177 PAVNQIETHPYFQRCCLVAFCKQKHGILVTAHTPLGGSTANTWFWGVSCLDDPVIKSLAE 236
Qy 245 KGNNTLAQVLIAGLRRGYVVLKSSNPKRIESNFKSIE--LSDADFEAINAVAKGRHFR 302
Db 237 KYGKTPAQVLIARWGLQRNTVPIPKYSKEERLQENFAVDFDFALISDEDMEKRSI--DRKYR 294
Qy 303 FVNMKDTFGYDVM 315
Db 295 TNQPAKFWGIDLF 307

RESULT 13
US-09-702-705-340
Query Match 28.6%; Score 494.5; DB 3; Length 308;
Best Local Similarity 36.4%; Pred. No. 1.3e-44; Mismatches 114; Conservative 61; Indels 15; Gaps 8;
Matches 114;
Qy 9 LNSGVKIPGVGFGTFASEGSGKETVAVTTALKTGYRHLDCAWYILNEGEVGEIGRDLFLK 68
Db 4 LNSGHTMPSVGLGVWEMDSP--AIRDLHLSALRIGYRHFDCADYQNEAEVGDALAEAF- 60
Qy 69 ENPSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYVDMFLVHWPAAEKNGQGE- 127
Db 61 QTLGVKREDLFTITKLWNSDHGH--VVEACKOSLKLRLDYLDLYLIHPVATRTGTGT 118
Qy 128 -EPKIGPDGKXVILKDLTENPEPTWRAMEKIVYEDRKARSIGVSNWTIADLEKMSKFAKVM 186
Db 119 TASALGDDG--VLDIDTTISLETTHAMEDIVSMGLVRSIGISNYDIFLTRCLAYAKIK 176
Qy 187 PHANQIEIHPPLNEELVOYCFKSNIMPVAYSPGSONQVPTTGERVS--ENKTLNEIAE 244
Db 177 PAVNQIETHPYFQRCCLVAFCKQKHGILVTAHTPLGGSTANTWFWGVSCLDDPVIKSLAE 236
Qy 245 KGNNTLAQVLIAGLRRGYVVLKSSNPKRIESNFKSIE--LSDADFEAINAVAKGRHFR 302
Db 237 KYGKTPAQVLIARWGLQRNTVPIPKYSKEERLQENFAVDFDFALISDEDMEKRSI--DRKYR 294
Qy 303 FVNMKDTFGYDVM 315
Db 295 TNQPAKFWGIDLF 307

RESULT 13
US-09-702-705-340
Query Match 28.6%; Score 494.5; DB 3; Length 308;
Best Local Similarity 36.4%; Pred. No. 1.3e-44; Mismatches 114; Conservative 61; Indels 15; Gaps 8;
Matches 114;
Qy 9 LNSGVKIPGVGFGTFASEGSGKETVAVTTALKTGYRHLDCAWYILNEGEVGEIGRDLFLK 68
Db 4 LNSGHTMPSVGLGVWEMDSP--AIRDLHLSALRIGYRHFDCADYQNEAEVGDALAEAF- 60
Qy 69 ENPSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYVDMFLVHWPAAEKNGQGE- 127
Db 61 QTLGVKREDLFTITKLWNSDHGH--VVEACKOSLKLRLDYLDLYLIHPVATRTGTGT 118
Qy 128 -EPKIGPDGKXVILKDLTENPEPTWRAMEKIVYEDRKARSIGVSNWTIADLEKMSKFAKVM 186
Db 119 TASALGDDG--VLDIDTTISLETTHAMEDIVSMGLVRSIGISNYDIFLTRCLAYAKIK 176
Qy 187 PHANQIEIHPPLNEELVOYCFKSNIMPVAYSPGSONQVPTTGERVS--ENKTLNEIAE 244
Db 177 PAVNQIETHPYFQRCCLVAFCKQKHGILVTAHTPLGGSTANTWFWGVSCLDDPVIKSLAE 236
Qy 245 KGNNTLAQVLIAGLRRGYVVLKSSNPKRIESNFKSIE--LSDADFEAINAVAKGRHFR 302
Db 237 KYGKTPAQVLIARWGLQRNTVPIPKYSKEERLQENFAVDFDFALISDEDMEKRSI--DRKYR 294
Qy 303 FVNMKDTFGYDVM 315
Db 295 TNQPAKFWGIDLF 307

Best Local Similarity 39.4%; Pred. No. 4e-45;
Matches 121; Conservative 51; Mismatches 106; Indels 29; Gaps 10;

QY 9 LNSGVKIPGVGTFPASEGSGKETVAVTTALKTGYRHLDCAWYLNNEGEVGEIGIRD 65
Db 3 TTVLSSGVEMPVIGLGLWRLE--KDELKEVILNAIKGYRHFDCAAHYKSEADVGEALAE 60
QY 66 FLKENPSVKREDIFVCTKVNHLHRYEDVLSIDSLKRLGLDLYVDMPLVHWPFAAEKNG 125
Db 61 AFKTG-LVKREBELFTTKIMNSDGH--VBEACKNSLEKLIQIDYLDLYLVHVPMPTRHNA 117
QY 126 QGEPK--TGPCKVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMSKEA 183
Db 118 IGTASLGED--KVLDIDVTISLQOTWEGEKTUSSLVGLVRSIGLSVEFLTRDCIAYS 175
QY 184 KVMPHANOIEIHPFLNPEELVOYCFCKNIMPVAYSPLGSONQVPTTGERSV--ENKTLNE 241
Db 176 KIKPAVSQFETHPYFQFSDLSVKFCMKHGVLPATHTPLGGAANKDMFGSVSPLDDPVLND 235
QY 242 IAEKGNVLAQVLIANGLRGCVVLPKSSNPKRIESNFKSIE--LSDADFEAINAVAKGR 299
Db 236 VAKYKGSVAQICLRWGIQRTAVIPKSSKIQRLENLEVLFEFOLSDCEDWQLIYSI--DR 293
QY 300 HRFVNMKDTFGYDW 315
Db 294 KYRTSLPKXTWGLDLY 309

RESULT 9
US-08-801-344-4
Sequence 4, Application US/08801344
Patent No. 6087140
GENERAL INFORMATION:
APPLICANT: Cameron, Douglas C.
APPLICANT: Shaw, Anita J.
APPLICANT: Altaras, Nedim E.
TITLE OF INVENTION: MICROBIAL PRODUCTION OF 1,2-PROPANEDIOL
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dewitt Ross & Stevens S.C.
STREET: 8000 Excelsior Drive, Suite 401
CITY: Madison
STATE: WI
COUNTRY: U.S.A.
ZIP: 53717-1914
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/801.344
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sara, Charles S.
REFERENCE/DOCKET NUMBER: 09820.037
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-831-2100
TELEFAX: 608-831-2106
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 316 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Rat lens aldose reductase

Query Match 28.9%; Score 499.5; DB 3; Length 316;

RESULT 10
US-09-498-599-4
Sequence 4, Application US/09498599
Patent No. 6303352
GENERAL INFORMATION:
APPLICANT: Cameron, Douglas C.
APPLICANT: Shaw, Anita J.
APPLICANT: Altaras, Nedim E.
TITLE OF INVENTION: MICROBIAL PRODUCTION OF
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dewitt Ross & Stevens S.C.
STREET: 8000 Excelsior Drive, Suite 401
CITY: Madison
STATE: WI
COUNTRY: U.S.A.
ZIP: 53717-1914
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/498,599
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sara, Charles S.
REFERENCE/DOCKET NUMBER: 09820.037
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-831-2100
TELEFAX: 608-831-2106
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 316 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Rat lens aldose reductase

US-09-498-599-4

RESULT 7

US-08-731-320B-4
; Sequence 4, Application US/08731320B
; Patent No. 6416985
; GENERAL INFORMATION:
; APPLICANT: Wayne D. Loescher,
; APPLICANT: John D. Everard
; APPLICANT: Rebecca Grumet
; TITLE OF INVENTION: DNA Encoding Mannose
; TITLE OF INVENTION: 6-Phosphate Reductase
; TITLE OF INVENTION: and Recombinants
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway
; CITY: Okemos
; STATE: Michigan
; COUNTRY: USA
; ZIP: 48864
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 5.25 inch,
; MEDIUM TYPE: 360 kb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731.320B
; FILING DATE: October 15, 1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ian C. McLeod
; REGISTRATION NUMBER: 20,931
; REFERENCE/DOCKET NUMBER: MSU 4.1-275
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (517) 347-4100
; TELEFAX: (517) 347-4103
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 310
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE:
; DESCRIPTION:
; HYPOTHETICAL: No
; ANTI-SENSE: No
; ORIGINAL SOURCE:
; ORGANISM: apple
; IMMEDIATE SOURCE:
; LIBRARY: N/A
US-08-731-320B-4

Query Match 29.0%; Score 501.5; DB 4; Length 310;

Best Local Similarity 34.8%; Pred. No. 2.3e-45;
Matches 110; Conservative 73; Mismatches 118; Indels 15; Gaps 8;

QY 6 TFTLSNGVKIPGVGFGTSESGKGTAVTTALKTYGRHLDCAWYVYLNAGEYGEIRD 65
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 3 TWTLSGVEPVIUGLWRLB--KDELKEVILNAIKIGYRHFDCAAHYKSEADVGEALAE 60
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 66 FLKENPVSFKREDIFVCTKWVHLRYEDVLSIDSLKRLGLDVDMFLVHPWPAEKNG 125
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 61 AFKGTG-LVKBRELEFITTITWNSDGH--VWEACKNSLEKLIQIDYLDLYLVHPMPTRHNA 117
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 126 QGEPK--IGPGKVILKDLTENPEPTWRAMEKIVEDRKARSIGVSNWTIADLEKMKKFA 183
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 118 IGTKASLLGED--KVLIDIVTISLQQTWEGVEKTVSLGLVRSIGLSNYELPLTRDCLAYS 175
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :

QY 184 KVMPHANOIEIHFFLPNEELVQYCFSKNIMPVAYSPLGSONQVPTTGBRVS--ENKTLNE 241
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 176 KIRPAVSQFETHPYFQRDSLVRFCMKHGVLPFAHTPLGGAANKDMFGSVPLDDPVLND 235
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 242 IAEKGCNTLAQVLIAMGLRGVYVLPKSNPKRIESNFKSIB--LSDADFEAINAVAKGR 299
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 236 VAKYKGKVAQICLRWGIQRTAVIPKSKIQRLKENLEVLFQSLSDQMQLIYSI--DR 293
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 300 HRFVNMKDTFGYDWN 315
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 294 KYRTSLPSKWTGLDVI 309
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :

RESULT 8

US-09-166-438-4
; Sequence 4, Application US/09166438
; Patent No. 6495317
; GENERAL INFORMATION:
; APPLICANT: Wayne D. Loescher,
; APPLICANT: John D. Everard
; APPLICANT: Rebecca Grumet
; TITLE OF INVENTION: DNA Encoding Mannose
; TITLE OF INVENTION: 6-Phosphate Reductase
; TITLE OF INVENTION: and Recombinants
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway
; CITY: Okemos
; STATE: Michigan
; COUNTRY: USA
; ZIP: 48864
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 5.25 inch,
; MEDIUM TYPE: 360 kb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/166,438
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/731,320
; FILING DATE: October 15, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ian C. McLeod
; REGISTRATION NUMBER: 20,931
; REFERENCE/DOCKET NUMBER: MSU 4.1-275
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (517) 347-4100
; TELEFAX: (517) 347-4103
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 310
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE:
; DESCRIPTION:
; HYPOTHETICAL: No
; ANTI-SENSE: No
; ORIGINAL SOURCE:
; ORGANISM: apple
; IMMEDIATE SOURCE:
; LIBRARY: N/A
US-09-166-438-4

Query Match 29.0%; Score 501.5; DB 4; Length 310;

Best Local Similarity 34.8%; Pred. No. 2.3e-45;
Matches 110; Conservative 73; Mismatches 118; Indels 15; Gaps 8;

Db 237 KYGKTPAQLVRLWGLQRTDWTWPKTSKVERLQENFDVFGFDISGDMRMKAI--DRKYR 294
QY 303 FVNMKDTFGYDVW 315
Db 295 TNQPAKFWGIDLY 307

RESULT 5
US-09-347-803-14
; Sequence 14, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
; APPLICANT: Farnodu, Layo O.
; APPLICANT: Hitz, Bill
; APPLICANT: Kinney, Tony
; APPLICANT: Orozco, Buddy
; TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes

; FILE REFERENCE: BB-1176
; CURRENT APPLICATION NUMBER: US/09/347.803
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092.952
; EARLIER FILING DATE: July 15, 1998
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Glycine max
US-09-347-803-14

Query Match 29.4%; Score 509.5; DB 3; Length 309;
Best Local Similarity 37.6%; Pred. No. 3.2e-46;
Matches 118; Conservative 59; Mismatches 122; Indels 15; Gaps 8;

QY 8 TLSNGVKIPGVGFGTFASGSGKGETYTAVTALKTGYRHLDCAWYILNEGEVSGIRDFL 67
Db 4 TLNNGFKPIGLGWRNEG--EIRDILNSIKIGYHFDCADYKNEAEVGDALKEAF 61
QY 68 KENPSVKREDIFVCTKYVNNHLHRYEDVLSDDSLKRLGIDYDMFLVHPPIAAEKNGQG 127
Db 62 -DSGLVKREDIFITTKLWNSDQGH--VLEACKDSLKLQILTYLDLYLVHFPVAVRHTGVG 118
QY 128 EPK--IGPDGKYVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMSKZAKV 185
Db 119 NTSSPLGDG--VLDDITTSLETTWHAEDLVSSGLVRSIGISNYDIFLTRDCLAYSKI 176
QY 186 MPANQIEIHPFLPNEELVQYCFSKNIMPVAYSPGSONQVPTTGERVS--ENKTLNEIA 243
Db 177 KPANVQIETHPYFQDSLVKFCQKHGICVTAHTPLGGAANAENWFGTVSCLDQVLKGLA 236
QY 244 EKGNTLAQVLIANGRLRGYVLPKSSNPKEIESNFK--STELSDADFEAINAVAKGRHF 301
Db 237 EYKKTAAQISLRWGIQRTVTVIPKSKLERLKENFQVDFELSKMEDMELIGSI--DRKY 294
QY 302 RFVNMKDTFGYDVW 315
Db 295 RTNQPAVFWGIDLY 308

RESULT 6
US-09-166-412-4
; Sequence 4, Application US/09166412
; Patent No. 6133504
; GENERAL INFORMATION:
; APPLICANT: Wayne D. Loescher,
; APPLICANT: John D. Everard
; APPLICANT: Rebecca Grumet
; TITLE OF INVENTION: DNA Encoding Mannose
; TITLE OF INVENTION: 6-Phosphate Reductase
; TITLE OF INVENTION: and Recombinants
; TITLE OF INVENTION: Produced Therefrom
; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway
; CITY: Okemos
; STATE: Michigan
; COUNTRY: USA
; ZIP: 48864
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 5.25 inch,
; MEDIUM TYPE: 360 kb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; OPERATING SYSTEM: (version 3.3)
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/166,412
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/731,320
; FILING DATE: October 15, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ian C. McLeod
; REGISTRATION NUMBER: 20,931
; REFERENCE/DOCKET NUMBER: MSU 4.1-275
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (517) 347-4100
; TELEFAX: (517) 347-4103
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 310
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE:
; DESCRIPTION:
; HYPOTHETICAL: No
; ANTI-SENSE: No
; ORIGINAL SOURCE:
; ORGANISM: apple
; IMMEDIATE SOURCE:
; LIBRARY: N/A
; US-09-166-412-4
Query Match 29.0%; Score 501.5; DB 3; Length 310;
Best Local Similarity 34.8%; Pred. No. 2.3e-45;
Matches 110; Conservative 73; Mismatches 118; Indels 15; Gaps 8;
QY 6 TPTLSNGVKIPGVGFGTFASGSGKGETYTAVTALKTGYRHLDCAWYILNEGEVSGIRD 65
Db 3 TPTLSNGVEMPVIGLGLWRLE--KDELKEVILNAIKIGYHFDCAAHYKSEADVGELAE 60
QY 66 FLKENPSVKREDIFVCTKYVNNHLHRYEDVLSDDSLKRLGIDYDMFLVHPPIAAEKNG 125
Db 61 AFKGTG-LVKGRELFITTKLWNSDGH--VVEACKSLKQLQIDYLDLYLVHYPMTKINA 117
QY 126 QGEPEK--IGPDGKYVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMSKPA 183
Db 118 IGTASLLGED--KVLDIDVTISLQQTWEGMEKTVSLGLVRSIGLSNVELFLTRDCLAYS 175
QY 184 KVMPHANQIEIHPFLPNEELVQYCFSKNIMPVAYSPGSONQVPTTGERVS--ENKTLNE 241
Db 176 KIKPAVSQFETHPYFQDSLVKFCMKHGVLPFAHTPLGGAANKDMFGSVSLDDPVLND 235
QY 242 IAEKGGNTLAQVLIANGRLRGYVLPKSSNPKEIESNFKSI--LSDADFEAINAVAKGR 299
Db 236 VAKYKGSVAQICLRWGIQRTAVIPKSKIQRLKENLEVLFLQSLSDQMIIYSI--DR 293
QY 300 HRFVNMKDTFGYDVW 315
Db 294 KYRTSLPSKWTGLDLY 309

```

; Sequence 6, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Hitz, Bill
; APPLICANT: Kinney, Tony
; APPLICANT: Kinney, Tony
; APPLICANT: Kinney, Tony
; TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
; FILE REFERENCE: BB-1176
; CURRENT APPLICATION NUMBER: US/09/347,803
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,952
; EARLIER FILING DATE: July 15, 1998
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Glycine max
US-09-347-803-6

Query Match      31.4%; Score 543.5; DB 3; Length 313;
Best Local Similarity 37.5%; Pred. No. 7.2e-50;
Matches 123; Conservative 62; Mismatches 108; Indels 35; Gaps 9;

QY 5 KTFTLSNGVKIPGVGFGTFASEGSKGETYTAVTALKTGYRHLDCAWYLYNEGEVGEGR 64
DB 6 KPFELNTGAKIPSVGLGTWQAE--PGVAVKAVTTAILVGRSHIDCAQAYNNQAEIGSALK 63

QY 65 DFLKENPSVKREDIFVCTKVNNHLHRYEDVLSIDDSLKRLGLDYVDMFLVHVAEKN 124
DB 64 K-LFDDGVVGRREDLWITSKWCSHDSEDPVKALDKTLQDLQDLVLYLHWPVMSK 122

QY 125 GQGEPIKIGPDGKYVILKDLTENPE--PTWRAMEKIYEDRKARSIGVSNWTIADLEKMSK 182
DB 123 SVG-----FKKEYLQDPDIPSTWKAMEALYDSGKARA.GVSNFSSKKLQDLWNI 171

QY 183 AKWPHANQIIEHPFLNEELVOYCFSKNIMPVAYSPGSGNQVPTTGERSVENKTLNEI 242
DB 172 ARVPPAVNQVELHPGQQAKLHAFCEKSGVHLSGYSPLGSGVGLKSD---ILKNPVVIEI 228

QY 243 AEKGGNTLAQVLIANGLRGVVLPKSSNPKRIESNFKSIELSDAD-----FEAI--NAV 295
DB 229 AEKLGKTPAQVALRWGLQGHVLPKSTNEARIKNLVDVYDWSIPEDLFPKFSKIQDL 288

QY 296 AKGRHFRFVNMKDTFG-----YDVWPEE 318
DB 289 IKGTFF---VDITYGAFKTVLELMDGE 312

```

```

RESULT 3
US-09-347-803-25
; Sequence 25, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Hitz, Bill
; APPLICANT: Kinney, Tony
; APPLICANT: Kinney, Tony
; APPLICANT: Kinney, Tony
; TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
; FILE REFERENCE: BB-1176
; CURRENT APPLICATION NUMBER: US/09/347,803
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,952
; EARLIER FILING DATE: July 15, 1998
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 25
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Medicago sativa
US-09-347-803-25

```

```

Query Match      30.2%; Score 523.5; DB 3; Length 313;
Best Local Similarity 37.2%; Pred. No. 1e-47;
Matches 123; Conservative 61; Mismatches 106; Indels 41; Gaps 10;

QY 5 KTFTLSNGVKIPGVGFGTFASEGSKGETYTAVTALKTGYRHLDCAWYLYNEGEVGEGR 64
DB 6 KPFQNTGAKIPSVGLGTWQAE--PGVAVKAVTTAVQVGRHIDCAEAYKNQSEIGSALK 63

QY 65 DFLKENPSVKREDIFVCTKVNNHLHRYEDVLSIDDSLKRLGLDYVDMFLVHVAEKN 124
DB 64 K-LCEDGVVGRBELWITSKWCSHDSEDPVKALDKTLQDLQDLVLYLHWPVMSK 121

QY 125 GQGEPIKIGPDGKYVILKDLTENPE--PTWRAMEKIYEDRKARSIGVSNWTIADLEKMSK 179
DB 122 GTGE-----FMGENLDHADIPSTWKALGALYDSGKAKAIGVSNFSTKKLQDL 168

QY 180 SKPAKVMPHANQIIEHPFLNEELVOYCFSKNIMPVAYSPGSGNQVPTTGERSVENKTL 239
DB 169 LDVARVPPAVNQVELHPGQQAKLHAFCEKSGVHLSGYSPLGSGVGLKSD---ILKNPVV 225

QY 240 NEIAEKGNTLAQVLIANGLRGVVLPKSSNPKRIESNFKSIELSDAD-----FEAI-- 292
DB 226 KEIAEKLGTPOVALRWGLQGHVLPKSTNEARIKNLVDVYDWSIPEDLFPKFSKIQ 285

QY 293 NAVAKGRHFRFVNMKDTFG-----YDVWPEE 318
DB 286 DKLIKGTFF---VNDITYGAFRTIBELMDGE 312

```

```

RESULT 4
US-09-347-803-10
; Sequence 10, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Hitz, Bill
; APPLICANT: Kinney, Tony
; APPLICANT: Kinney, Tony
; APPLICANT: Kinney, Tony
; TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
; FILE REFERENCE: BB-1176
; CURRENT APPLICATION NUMBER: US/09/347,803
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,952
; EARLIER FILING DATE: July 15, 1998
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Zea mays
US-09-347-803-10

Query Match      30.0%; Score 518.5; DB 3; Length 308;
Best Local Similarity 37.7%; Pred. No. 3.4e-47;
Matches 118; Conservative 60; Mismatches 120; Indels 15; Gaps 8;

QY 9 LSNQVKIPGVGFGTFASEGSKGETYTAVTALKTGYRHLDCAWYLYNEGEVGEGRDFLK 68
DB 4 LSSGHRMPAVGLGVWRME--KADIRGLIHTALURVGRHLDCAADYQNEAEVGDALAEAF- 60

QY 69 ENPSVKREDIFVCTKVNNHLHRYEDVLSIDDSLKRLGLDYVDMFLVHVAEKNQGO- 127
DB 61 QTGLVKREDLFTITTKLWNSDHGH--VLEACKDSKKLQDLVLYLHFPVATRHGTGVT 118

QY 128 -EPKIGPDGKYVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMSKFAKM 186
DB 119 TSSALGDDG--VLDIDTTISLETTHAMEELVSMGLVSGISNISNYDIFLTRCLAYATIK 176

QY 187 PHANQIIEHPFLNEELVOYCFSKNIMPVAYSPGSGNQVPTTGERS--ENKTLNEIAE 244
DB 177 PAVNQIETHPTFQRDSLVKFCCKHGICVTAHTPLGSGTANAEMFGTVSCLDDBVIKSLAD 236

QY 245 KGNTLAQVLIANGLRGVVLPKSSNPKRIESNFK--SIELSDADFEAINAVAKGRHFR 302

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2004, 02:50:06 ; Search time 45 Seconds
(without alignments)
372.854 Million cell updates/sec

Title: US-10-004-115B-1

Perfect score: 1731

Sequence: 1 MSNGKTTLSNGVKGPGVF.....MKDTFGVDVWPEETAKNLSA 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pdp.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pdp.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pdp.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pdp.*
- 5: /cgn2_6/ptodata/2/iaa/PTCUS_COMB.pdp.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	547.5	31.6	344	4	US-09-630-983A-20
2	543.5	31.4	313	3	US-09-347-803-6
3	523.5	30.2	313	3	US-09-347-803-25
4	518.5	30.0	308	3	US-09-347-803-10
5	509.5	29.4	309	3	US-09-347-803-14
6	501.5	29.0	310	3	US-09-166-412-4
7	501.5	28.0	310	4	US-08-731-320B-4
8	501.5	28.0	310	4	US-09-166-438-4
9	499.5	28.9	316	3	US-08-801-344-4
10	499.5	28.9	316	4	US-09-498-599-4
11	498	28.8	298	4	US-09-134-000C-6059
12	494.5	28.6	308	3	US-09-347-803-12
13	487	28.1	316	4	US-09-702-705-340
14	487	28.1	316	4	US-09-736-457-340
15	487	28.1	316	4	US-09-614-124B-340
16	487	28.1	316	4	US-09-589-184-340
17	487	28.1	316	4	US-09-166-412-2
18	486.5	28.1	309	3	US-09-347-803-26
19	486.5	28.1	309	3	US-08-731-320B-2
20	486.5	28.1	309	4	US-09-166-438-2
21	486.5	28.1	309	4	US-09-166-438-2
22	482.5	27.9	308	3	US-09-347-803-8
23	472.5	27.3	316	1	US-08-585-595-3
24	466	26.9	291	4	US-09-107-532A-5063
25	456	26.3	323	4	US-09-702-705-1821
26	456	26.3	323	4	US-09-736-457-1821
27	456	26.3	323	4	US-09-671-325-1821

RESULT 1
US-09-630-983A-20
; Sequence 20, Application US/09630983A
; Patent No. 6630330
; GENERAL INFORMATION:
; APPLICANT: Porro, Danilo
; APPLICANT: Sauer, Michael
; TITLE OF INVENTION: Ascorbic Acid Production from Yeast
; FILE REFERENCE: 2028.594000
; CURRENT APPLICATION NUMBER: US/09/630,983A
; CURRENT FILING DATE: 2000-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 20
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-630-983A-20

Query Match 31.6%; Score 547.5; DB 4; Length 344;
Best Local Similarity 37.4%; Pred. No. 3.1e-50;
Matches 122; Conservative 58; Mismatches 91; Indels 55; Gaps 8;

QY	7	FTLSNGVKIPGVGFGTFASEGSKEGTYTAVTTALKTGYRHLDCAWYILNEGVGEGIDF	66
DB	25	FTLSNGVRIPALGLGTANPHEKLAETKQAKAAIKAGYRHIDTANAYETEPVGAIKEL	84
QY	67	LKENPSVKREDIFVCTKVMNHLHYEDVLW-----SIDSLKELGLDYVDMFLVHWPIAA	121
DB	85	L-EDGSIKREDLFIITKVM-----PVLWDEVDRLNESLKALGLEVYVLLLOHWPCLCF	136
QY	122	EK-----NCGEPKGPCKYVILKDLTENPEPTWAMEKIY---EDRKA	163
DB	137	EKIDPKGISGLVKVTEVDSDGKTMVAADGYL-----ETKQLEKIYLDPNDRHV	186
QY	164	RSIGVSNWTIADLEKMSKPAKVMPHANQIEIHPFLPNEELVOYCFSKNIMPVAYPLGSG	223
DB	187	RAIGVSNFISIEYLERLIKECRVKPTVQVETHFPLPQMELRKFCFMHDILLTAYSPGLSH	246
QY	224	N-----OVPTTGRVSEBNTLNIAEKGNTLAQVLIANGREGYVVLPKSSNPKEISNF	279
DB	247	GAPNLKIP-----LVKKLAEKYNTGNLLIYSHIROGTIVIPRSLNPFVRISSI	296
QY	280	KSEIISDADFEAINAVAKGRHFRFVN	305
DB	297	EPASUTKDELQELNDPGEKYVPRFID	322

RESULT 2

US-09-347-803-6

ALIGNMENTS


```
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1473
; LENGTH: 344
; TYPE: PRI
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1473

Query Match      31.6%; Score 547.5; DB 15; Length 344;
Best Local Similarity 37.4%; Pred. No. 1.3e-47;
Matches 122; Conservative 58; Mismatches 91; Indels 55; Gaps 8;

QY 7 FTLSNGVKIPGVGFTFASGSGEYTYATTALKTGYRHLDCAWYYLNEGEVGEIGIRDF 66
Db 25 FSLNNGVRIIPALGLGTANPHEKLAETKQAVKAAKAGYRHIDTAWAYETEPFVGEAIKEL 84

QY 67 LKENPSVKREDIFVCTKVNHLHRYEDVLW-----SIDDSLKELGLDYVDMELVHWPIAA 121
Db 85 L-EDGSIKREDLFTTTKW-----PVLMDVDRSLNESLKALGLEYYDLLLLQHWPLCF 136

QY 122 EK-----NGQGEPKIGPDGKYVILKDLTENPEPTWRAMEKIY---EDRKA 163
Db 137 EKIKDPKGISGLVKTPVDDSGKTWYAADGDL-----ETYKQLEKIYLDPNDRV 186

QY 164 RSTGVSNWTIADLEKSKFAKMPHANOIEHPFLNEELVQYCFSKNIMPVAYSPLGSO 223
Db 187 RAIGVSNFSIEYLERLIKECRVKPTVNOVETHPHLPQOMELRKFCFMDHILLTAYSPLGSH 246

QY 224 N---QVPTTGERVSENKTLNEIAEKGNTLAQVLIANGLRGYYVLPKSSNPKRIESNF 279
Db 247 GAPNLKIP-----LVKKLAEKYNTVGNLLISYHROGTIVIPRSLNPVRISSSI 296

QY 280 KSIELSDADFEAINAVAKGRHFRPN 305
Db 297 EFASLTQDELQELNDFGEKYPVRFD 322
```

Search completed: May 28, 2004, 14:07:19
Job time : 49 secs

QY	5	KYFTLSNGVKIPGVGFTFASBGSGKGYTYTAVTALTKTGYRHLDCAYNYLNEGEVBEGIR	64
Db	11	KYFFLSNGBQIPALGLGTWS--SPQVVGQAVEQALDLGYRHLDCAAIYNGEAEIGATLA	68
QY	65	DFLKENPFSVKREDIIFVCTKVMNHLHRYEDVLWS:DDSLKRLGLDYDMFLVHVPPIAAEKN	124
Db	69	NAFTKG-VVKREELMITSKLSNAHHPDAVLPAALEKTLQDLGLDYLDLYLHHPVVI---	124
QY	125	GQGEPIG-PDGKYVILKDLTENPEPTWRAMEKIYEDRKARSICVSNWTTADLEKSKFA	183
Db	125	---QPDVGPESSGDQLLPTPASLEGTWQALEKAVDGLGCHIGVSNFSUKLEWLMSLA	181
QY	184	KYMPHANKIIEHPFLNBLVQYCSKNIMPVAYSPLGSCNQVPTTCERVSSENKTL----	239
Db	182	RIPPAVNOVELHPYLQOSDLLTFANSONILLTAYSPLGSDR-PAAPQQAEPKLLTDPV	240
QY	240	-NEIAEKGNTLAAOVLIANGLRGGVYVLPKSNPKRIESNFKSLELSDADFEAINAVAKG	298
Db	241	INGIAAEQCSAAQVLLAWAITQGGVTIPTKSNPNERLEQNLRADITLTDSEMAKIALLD	300
QY	299	RHREFNV 305	
Db	301	RHYRYVS 307	

RESULT 13

```

US-10-369-493-2292
/ Sequence 2292, Application US/10369493
/ Publication No. US2003023675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ CURRENT FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 2292
/ LENGTH: 315
/ TYPE: prt
/ ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2292

```

```

Query Match      31.7%; Score 548.5; DB 15; Length 315;
Best Local Similarity 40.5%; Pred. No. 8.9e-48;
Matches 130; Conservative 52; Mismatches 102; Indels 37; Gaps 8;

QY      7 FTLSNGVXIPGVGFGFTASEGSKGETYATVTTAKTGYRHLCDAWYILNEGEVGEGRD 66
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      12 FTLPNGDXIPSIGLGWTWS--GKDETNAVCALKAGYRHDITAIYGNEXEIGEGR- 68

QY      67 LKENSVKREDIFVCTKVWHLHRYEDVLSIDSLKRLGLDYVDMFLVHPHPIA----- 120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      69 -----SGVPRTDIWTWSKLWCNAHRAGVLPALAEKTIQDLNLEYIATYIHPFALLSGPE 124

QY      121 -AEKNQGOEPIKPGDKYVILKOLTENPEFTWAMEKIYEDRKARSIGVSNWTIADLEKM 179
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      125 ELPRNEKGE-----LIYEDVP--IETWQAMEELLETGKVRYIGISNPFNNEYLD 173

QY      180 SKFAKVMPHANOIETHPLPNEELVOYCFPSKNIMPVAYSLGSONQVPTTG-ERVSENKT 238
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      174 LKIAKVKTIIHQMLEHLPVOTVELEKHKLIQHVSAYSPLANQDANYNSOISKLIHEHT 233

QY      239 LNEIAEKGGN--TTAQLVIAWLRGVRVYLPKSSNPKRIESNPKSIEISDADFPAINAVA 296
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      234 LVDIANARAGEGITTANIAISWAKRGTSVLPKGVNSRVSFNFLYPIPTDKEMEAINNIG 293

```

```

QY      297 KGRHF-----RFVANKD 308
        |||
        |||.
Db      294 VVRFSHGKFAKPWFVGLOD 314
        |||
        |||.

RESULT 14
US-10-424-599-282510
; Sequence 282510, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovacic David X
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other
; TITLE OF INVENTION: Plants and Uses Thereof for Plant
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 282510
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_97128C.1.pep
US-10-424-599-282510

```

Query Match	31.6%	Score	547.5	DB 12	Length	313			
Best Local Similarity	37.2%	Pred. No.	1.1e-47						
Matches	121	Conservative	66	Mismatches	109	Indels	29	Gaps	8

QY	5	KTFILSNVKVTPGVGTFPASEGSGKETTAVTALTKTGYRHLDCAWYVNLNBEVGEVGEIR	64
DB	6	KFFELNTGAKIPISGLGTWQAE--PGVVAELTAITQVGYRHIDCASAYKNQAEIGSALK	63
QY	65	DFLKENPSVKREDIFVCTKYVNEHLHRYEDVLMISIDSLKRLGLDLYVDFPLVHPVIAAEKN	124
DB	64	K-LFDDGVVKEEDLWITSKLWCSDHAPEDVPKALDKTQLQELQDLYDLJYLINHPYMK--	120
QY	125	GQGBPKTGFQGYVILKDLTENPE--PTWRAMEKIYEDRKARSIGVSNWTIADLEKMGKF	182
DB	121	-----SGTFLNKLEYLEQPDINTWRAMEALYDSDKARAIGVSNPSSKKLQDFLDI	171
QY	193	AKVMPHANQIIEHFFLNELVQVCSKNIMPVAYSPLGSONQVPTTGERVSENKTLNEI	242
DB	172	ARVUPAVNQVELHPGWQOPKLRAPFCSKEIHLGSGYSPLGSPAAALKSD---ILKNPVVTEI	228
QY	243	AEKGNTIAQVLTAWGLRRGVVYLPKSSNPKRIESNFKSIELSDADFAINAVAKGRHFR	302
DB	229	AERLGKTPQAVLRWGLQAGHSVLPKSTNESRIKGNFDIFDWSIQ-DLMTKISIKQER	287
QY	303	FWN-----MKOTFG-----YDVWPEE	318
DB	288	LVKASFFVHETYGAYRSIEDFDWGE	312

```

RESULT 15
US-10-369-493-1473
; Sequence 1473. Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039

```



```
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 09/494,921
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.11
; SEQ ID NO 44
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic protein derived from Saccharomyces cerevisiae Gcy1 redu
; OTHER INFORMATION: ctase, having a glycine inserted at position 2 in the amino acid
; OTHER INFORMATION: sequence
US-09-734-237B-44

Query Match      33.2%; Score 574.5; DB 12; Length 313;
Best Local Similarity 40.8%; Pred. No. 1.8e-50;
Matches 127; Conservative 54; Mismatches 85; Indels 45; Gaps 7;

QY      5 KFTLSNGVKIPGVGFGTFASEGSGKETATVTTALKTGVRHLDCAWYVINEGEVGBGIR 64
DB      11 KILSNTGAQIPQIGLGTWQSK--ENDAYKAVLTALKOGYRHDTAAIYNEDQVGQAII 68
QY      65 DFLKENSVKREDIFVCTKWNHLYRYEDVLWSDSLKRLGLDYVDMFLVHWPPIAAEKY 124
DB      69 D-----SGVPRKEIFVTIKLWCTQHEPEV--ALDQSLKRLGLDYVDMFLVHWPPIAAEKY 118
QY      125 GQGEPKIGDQKYYILKDLTENP-----PTWEAKEIYEDRKARSIGV 168
DB      119 -----DPAITKNEIDILSVPTKDGSRADVITNNFTKMWELMOELPKTKRAVGV 169
QY      169 SNWTIADLEKM--SKFAKVMFPHANQIEIHPFLPNEELVQYCFKSNIMPVAYSPLGSONQV 226
DB      170 SNFSINNKLKLLASQGNKLTAPAAQVEIHPPLPQDELINCKSGIIVVEAYSPLGS---- 225
QY      227 PTTGERSVENKTLNEIAEKGNTIAQVLIAGLRRGVVLPKSNPKRIESNPKSIELSD 286
DB      226 --TDAPLKEVPIEIAKNNQVPGHVVISHWQGVVLPKSNVNPDRITNKRIFTLST 283
QY      287 ADFAINAVAK 297
DB      284 EDFAINNIISK 294

RESULT 8
US-10-369-493-2285
; Sequence 2285, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2285
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2285

Query Match      32.8%; Score 568; DB 15; Length 321;
Best Local Similarity 43.0%; Pred. No. 8.8e-50;
Matches 131; Conservative 59; Mismatches 89; Indels 26; Gaps 10;

QY      7 FTLSNGVKIPGVGFGTFASEGSGKETATVTTALKTGVRHLDCAWYVINEGEVGBGIRDF 66
```

```
DB      16 FTIADGSKIPGLGLGTWRSE--PNQTKNAVKTALQGYRHDAAAIYGNEDVGGDIKE- 72
QY      67 LKENSVKREDIFVCTKWNHLYRYEDVLWSDSLKRLGLDYVDMFLVHWPPIAAEKNGQ 126
DB      73 ----SGVPRKDIWVTSKLCWNAHAPEAVPKALEKTLKDLKLDYLDYLIHFPVSF-KTGE 127
QY      127 GEPKIGDQKYYILKDLTENP-EPTWRAVEKIYEDRKARSIGVSNWTIADLEKMSKFAKV 185
DB      128 DRFPKDKDGNLIYERK---NPIETWKAWEKLETKGVKVRHIGLSNFDNTNLERILKAVK 183
QY      186 MPAHQIEIHPFLPNEELVQYCFKSNIMPVAYSPLGSON-----CVPTTGERVSENKTLN 240
DB      184 KPAVQOMELHPFLPQTEFVEKHKGLGIHVAIYSPFGNQNTIYESKIP----KLIEHETIQ 239
QY      241 ETAE-KG-GNTLAQVLIAGLRRGVVLPKSNPKRIESNPKSIELSDADFAINAVAKG 298
DB      240 KIAKSGEGVGTATIAVSWAITRGTSVIPKSVNEQRIKSNPKYIPLTKEDMDINSI--G 297
QY      299 RHFRF 303
DB      298 IRARF 302

RESULT 9
US-09-801-368-438
; Sequence 438, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 438
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-438

Query Match      32.1%; Score 555; DB 9; Length 312;
Best Local Similarity 40.1%; Pred. No. 1.9e-48;
Matches 128; Conservative 60; Mismatches 97; Indels 34; Gaps 9;

QY      2 SNGKFTLSNGVKIPGVGFGTFASEGSGKETATVTTALKTGVRHLDCAWYVINEGEVGE 61
DB      7 NSSATLKLTNGASIPVLGFGTWSVDNNG--YHVSIAALKAGYRHIDAAAIYLAEEVEVGR 64
QY      62 GIRDFLKENPVSVKREDIEVCTKWNHLYRYEDVLWSDSLKRLGLDYVDMFLVHWPPIAA 121
DB      65 AIKD-----SGVPRKEIITTKLNGTEQR--DPEALANKSLKRLGLDYVDMFLVHWPPI 117
QY      122 EKNQGEKPIGDPKGKIVILKDLTE-----NPEPTWRAVEKIYEDRKARSIGVSNW 171
DB      118 KTRV-----TDGNVLGCIPTLEDGTIDTKEWNFIKTWELMQELPDKTKTGAVGNSF 171
```

; SEQ ID NO 3396

; LENGTH: 255

; TYPE: PRT

; ORGANISM: Neurospora crassa

US-10-369-493-3396

Query Match

Best Local Similarity 56.2%; Score 972.5; DB 15; Length 255;

Matches 174; Conservative 35; Mismatches 45; Indels 1; Gaps 1;

QY 16 PGVGFGRFASGSGGTYTAVTTALKTGYRHLDCAWYLYNEGEVGGIRDFLKENSVKR 75

Db 1 PAVGFGTFANAGAGETTYAAKVALEVGVRHLDCAWFYQNEDEVGQALAEFLNKHDKVR 60

QY 76 EDIFVCTKWNHLHRYEDVLSIDSLKRLGLDYVDMFLVHWPFAAEKN-GQGEPIKIGPD 134

Db 61 EDIFCTKWNHLHPEDEVKWSLQNSLDKLVVDVDFLIHWPFAAEKDEATNMPKIGPD 120

QY 135 GKVVILKDLTENPEPTWRAMEKIYEDKARSIGVSNWTIADLEKSKFAKVMPHANQIEI 194

Db 121 GXYIIRKELTENPEPTWRAMEDVDAGTKRSIGVSNWTIADLEKSKFAKVMPHANQIEI 180

QY 195 HPFLPNEELVQYCFKSNMFPVAYSPGSGNQVPTTGERSVSENKTLNEIAEKGGNTLAQVL 254

Db 181 HPFLPNEELVQYCFKSNMFPVAYSPGSGNQVPTTGERSVSENKTLNEIAEKGGNTLAQVL 240

QY 255 IAWGLRGRGVVLPKS 269

Db 241 LAWGLRGRGVVLPKS 255

RESULT 5

US-09-734-237B-42

; Sequence 42, Application US/09734237B

; Publication No. US20030064432A1

; GENERAL INFORMATION:

; APPLICANT: Rozzelli, J. David

; APPLICANT: Bui, Peter

; APPLICANT: Hua, Ling

; TITLE OF INVENTION: SYNTHETIC GENES FOR ENHANCED EXPRESSION

; FILE REFERENCE: B593:40608

; CURRENT APPLICATION NUMBER: US/09/734,237B

; CURRENT FILING DATE: 2000-12-08

; PRIOR APPLICATION NUMBER: 09/494,921

; PRIOR FILING DATE: 2000-01-31

; NUMBER OF SEQ ID NOS: 79

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 42

; LENGTH: 312

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-09-734-237B-42

Query Match

Best Local Similarity 33.2%; Score 574.5; DB 12; Length 312;

Matches 127; Conservative 54; Mismatches 85; Indels 45; Gaps 7;

QY 5 KTFTLSNGVKIPGVGFGTFASGSGGTYTAVTTALKTGYRHLDCAWYLYNEGEVGGIR 64

Db 10 KILSLNTGAQIPQIGLGTWQSK--ENDAYKAVLTALKDGYRHIDTAIYRNEDQVQGAIK 67

QY 65 DFLKENSVKREDIEVCTKWNHLHRYEDVLSIDSLKRLGLDYVDMFLVHWPFAAEKN 124

Db 68 D-----SGVPREEIFVTIKMCTQHEPEV--ALDQSLKRLGLDYVDMFLVHWPFAAEKN 117

QY 125 GQGEPIKIGPDGKYVILKDLTENPE-----PTWRAMEKIYEDRKARSIGV 168

Db 118 -----DPAYIKNEDILSVPTKDGSRVADITNNMFIKTWELMQELPKTKTKAVGV 168

QY 169 SNWTIADLEKM--SKFAKVMPHANQIEIHPFLPNEELVQYCFKSNMFPVAYSPGSGNQV 226

Db 169 SNFSINNLLKDLASQGNKLTTPAANQVETHPILLPODELINFCCKSGIIVVEAYSPLGS 224

QY 227 PTTGERSVSENKTLNEIAEKGGNTLAQVLIAWGLRGRGVVLPKSNPKRIEBSNFKSIELSD 286

Db 225 --TDAPLKEPVIIEIAKNNVQPGHVVISHWVQGVVLPKSNVNPDAIKTNRKIFTLST 282

QY 287 ADPEAINAVAK 297

Db 283 EDFAINNISK 293

RESULT 7

US-09-734-237B-44

; Sequence 44, Application US/09734237B

; Publication No. US20030064432A1

; GENERAL INFORMATION:

; APPLICANT: Rozzelli, J. David

; APPLICANT: Bui, Peter

; APPLICANT: Hua, Ling

; TITLE OF INVENTION: SYNTHETIC GENES FOR ENHANCED EXPRESSION

; FILE REFERENCE: B593:40608

; CURRENT APPLICATION NUMBER: US/09/734,237B

QY 227 PTTGERSVSENKTLNEIAEKGGNTLAQVLIAWGLRGRGVVLPKSNPKRIEBSNFKSIELSD 286

Db 225 --TDAPLKEPVIIEIAKNNVQPGHVVISHWVQGVVLPKSNVNPDAIKTNRKIFTLST 282

QY 287 ADPEAINAVAK 297

Db 283 EDFAINNISK 293

RESULT 6

US-10-369-493-22372

; Sequence 22372, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 22372

; LENGTH: 312

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-10-369-493-22372

Query Match

Best Local Similarity 33.2%; Score 574.5; DB 15; Length 312;

Matches 127; Conservative 54; Mismatches 85; Indels 45; Gaps 7;

QY 5 KTFTLSNGVKIPGVGFGTFASGSGGTYTAVTTALKTGYRHLDCAWYLYNEGEVGGIR 64

Db 10 KILSLNTGAQIPQIGLGTWQSK--ENDAYKAVLTALKDGYRHIDTAIYRNEDQVQGAIK 67

QY 65 DFLKENSVKREDIEVCTKWNHLHRYEDVLSIDSLKRLGLDYVDMFLVHWPFAAEKN 124

Db 68 D-----SGVPREEIFVTIKMCTQHEPEV--ALDQSLKRLGLDYVDMFLVHWPFAAEKN 117

QY 125 GQGEPIKIGPDGKYVILKDLTENPE-----PTWRAMEKIYEDRKARSIGV 168

Db 118 -----DPAYIKNEDILSVPTKDGSRVADITNNMFIKTWELMQELPKTKTKAVGV 168

QY 169 SNWTIADLEKM--SKFAKVMPHANQIEIHPFLPNEELVQYCFKSNMFPVAYSPGSGNQV 226

Db 169 SNFSINNLLKDLASQGNKLTTPAANQVETHPILLPODELINFCCKSGIIVVEAYSPLGS 224

QY 227 PTTGERSVSENKTLNEIAEKGGNTLAQVLIAWGLRGRGVVLPKSNPKRIEBSNFKSIELSD 286

Db 225 --TDAPLKEPVIIEIAKNNVQPGHVVISHWVQGVVLPKSNVNPDAIKTNRKIFTLST 282

QY 287 ADPEAINAVAK 297

Db 283 EDFAINNISK 293

QY 61 EGIRDFLENPSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYDMFLVHWPFA 120
DB 61 EGIRDFLENPSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYDMFLVHWPFA 120
QY 121 AEKNGQBEKIPGDPKGVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMS 180
DB 121 AEKNGQBEKIPGDPKGVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMS 180
QY 181 KFAKVMPHANQIEIHPFLPNEELVOYCFSKNIMPVAYSPGSONQVPTTGERVSENKTLN 240
DB 181 KFAKVMPHANQIEIHPFLPNEELVOYCFSKNIMPVAYSPGSONQVPTTGERVSENKTLN 240
QY 241 EIAKGGNTLAQVLIANGLRGYVLPKSNPKRIESNFKSIELSDADFEAINAVAKGRH 300
DB 241 EIAKGGNTLAQVLIANGLRGYVLPKSNPKRIESNFKSIELSDADFEAINAVAKGRH 300
QY 301 FRFVNKMTFGYDVPETAKNLSA 325
DB 301 FRFVNKMTFGYDVPETAKNLSA 325

RESULT 2

US-10-327-108-3

; Sequence 3, Application US/10327108

; Publication No. US20030186400A1

; GENERAL INFORMATION:

; APPLICANT: ASAKO, HIROYUKI

; APPLICANT: WAKITA, RYUHEI

; APPLICANT: ITOH, NOBUYA

; TITLE OF INVENTION: METHOD FOR PRODUCING OPTICALLY ACTIVE

; FILE REFERENCE: 073517

; CURRENT APPLICATION NUMBER: US/10/327,108

; CURRENT FILING DATE: 2002-12-24

; PRIOR FILING DATE: 2001-12-27

; PRIOR FILING DATE: 2001-12-27

; PRIOR FILING DATE: 2001-12-27

; PRIOR FILING DATE: 2002-04-10

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 3

; TYPE: PRT

; LENGTH: 325

; ORGANISM: Penicillium citrinum

; US-10-327-108-3

Query Match 100.0%; Score 1731; DB 14; Length 325;
Best Local Similarity 100.0%; Pred. No. 4.5e-170;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNGKTFILSGVKIPGVGFTFASGSGKGETTAVTTALTGTGRHLDCAWYVLYNEGEVG 60
DB 1 MSNGKTFILSGVKIPGVGFTFASGSGKGETTAVTTALTGTGRHLDCAWYVLYNEGEVG 60

QY 61 EGIRDFLENPSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYDMFLVHWPFA 120
DB 61 EGIRDFLENPSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYDMFLVHWPFA 120

QY 121 AEKNGQBEKIPGDPKGVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMS 180
DB 121 AEKNGQBEKIPGDPKGVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMS 180

QY 181 KFAKVMPHANQIEIHPFLPNEELVOYCFSKNIMPVAYSPGSONQVPTTGERVSENKTLN 240
DB 181 KFAKVMPHANQIEIHPFLPNEELVOYCFSKNIMPVAYSPGSONQVPTTGERVSENKTLN 240

QY 241 EIAKGGNTLAQVLIANGLRGYVLPKSNPKRIESNFKSIELSDADFEAINAVAKGRH 300
DB 241 EIAKGGNTLAQVLIANGLRGYVLPKSNPKRIESNFKSIELSDADFEAINAVAKGRH 300

QY 301 FRFVNKMTFGYDVPETAKNLSA 325
DB 301 FRFVNKMTFGYDVPETAKNLSA 325

RESULT 3

US-10-608-533-1

; Sequence 1, Application US/10608533

; Publication No. US20040091902A1

; GENERAL INFORMATION:

; APPLICANT: ASAKO, HIROYUKI

; APPLICANT: SHIMIZU, MASATOSHI

; TITLE OF INVENTION: MODIFIED REDUCTASE AND ITS GENE

; FILE REFERENCE: Q76265

; CURRENT APPLICATION NUMBER: US/10/608,533

; PRIOR FILING DATE: 2003-06-30

; PRIOR FILING DATE: 2002-07-02

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 1

; LENGTH: 325

; TYPE: PRT

; ORGANISM: Penicillium citrinum

; US-10-608-533-1

Query Match 100.0%; Score 1731; DB 16; Length 325;

Best Local Similarity 100.0%; Pred. No. 4.5e-170;

Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNGKTFILSGVKIPGVGFTFASGSGKGETTAVTTALTGTGRHLDCAWYVLYNEGEVG 60
DB 1 MSNGKTFILSGVKIPGVGFTFASGSGKGETTAVTTALTGTGRHLDCAWYVLYNEGEVG 60

QY 61 EGIRDFLENPSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYDMFLVHWPFA 120
DB 61 EGIRDFLENPSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYDMFLVHWPFA 120

QY 121 AEKNGQBEKIPGDPKGVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMS 180
DB 121 AEKNGQBEKIPGDPKGVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMS 180

QY 181 KFAKVMPHANQIEIHPFLPNEELVOYCFSKNIMPVAYSPGSONQVPTTGERVSENKTLN 240
DB 181 KFAKVMPHANQIEIHPFLPNEELVOYCFSKNIMPVAYSPGSONQVPTTGERVSENKTLN 240

QY 241 EIAKGGNTLAQVLIANGLRGYVLPKSNPKRIESNFKSIELSDADFEAINAVAKGRH 300
DB 241 EIAKGGNTLAQVLIANGLRGYVLPKSNPKRIESNFKSIELSDADFEAINAVAKGRH 300

QY 301 FRFVNKMTFGYDVPETAKNLSA 325
DB 301 FRFVNKMTFGYDVPETAKNLSA 325

RESULT 4

US-10-369-493-3396

; Sequence 3396, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10452052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; PRIOR FILING DATE: 2003-02-28

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 14:03:12 ; Search time 48 Seconds
(without alignments)
1890.752 Million cell updates/sec

Title: US-10-004-115B-1

Perfect score: 1731

Sequence: 1 MSNGKFTLLSNGVKIPGVGF.....MKDTFGYDWPBETAKNLISA 325

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1151071 seqs, 279249464 residues

Total number of hits satisfying chosen parameters: 1151071

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1731	100.0	325	14	US-10-004-115A-1
2	1731	100.0	325	14	US-10-327-108-3
3	1731	100.0	325	16	US-10-608-533-1
4	972.5	56.2	255	15	US-10-369-493-3396
5	574.5	33.2	312	12	US-09-734-237B-42
6	574.5	33.2	312	15	US-10-369-493-22372
7	574.5	33.2	313	12	US-09-734-237B-44
8	568	32.8	321	15	US-10-369-493-2285
9	555	32.1	312	9	US-09-801-368-438
10	555	32.1	312	15	US-10-369-493-1693
11	555	32.1	313	12	US-09-734-237B-39
12	553	31.7	327	15	US-10-369-493-2759
13	548.5	31.9	315	15	US-10-369-493-2292
14	547.5	31.6	313	12	US-10-424-599-282510
15	547.5	31.6	344	15	US-10-369-493-1473

16	547	31.6	328	13	US-10-040-416-6	Sequence 6, Appli
17	545.5	31.5	328	13	US-10-040-416-4	Sequence 4, Appli
18	545	31.5	315	12	US-10-424-599-183977	Sequence 183977,
19	543.5	31.4	313	10	US-09-882-691-6	Sequence 6, Appli
20	543.5	31.4	313	12	US-10-424-599-221378	Sequence 221378,
21	541	31.3	322	12	US-09-734-237B-32	Sequence 32, Appli
22	538	31.1	330	9	US-09-800-487A-2	Sequence 2, Appli
23	538	31.1	330	13	US-10-040-416-2	Sequence 221377,
24	534.5	30.9	313	12	US-10-424-599-221377	Sequence 16605, A
25	530	30.6	277	15	US-10-369-493-16605	Sequence 252130,
26	525.5	30.4	304	12	US-10-424-599-252130	Sequence 43149, A
27	525.5	30.4	305	12	US-10-425-114-43493	Sequence 252130,
28	525	30.3	279	12	US-10-282-122A-46356	Sequence 46356, A
29	524.5	30.3	274	15	US-10-369-493-16462	Sequence 16462, A
30	523.5	30.2	341	12	US-09-882-691-25	Sequence 25, Appli
31	523.5	30.2	341	12	US-10-425-114-48430	Sequence 48430, A
32	521	30.1	300	12	US-10-282-122A-52772	Sequence 52772, A
33	518.5	30.0	308	10	US-09-882-691-10	Sequence 10, Appli
34	518.5	30.0	339	12	US-10-425-114-59966	Sequence 59966, A
35	516	29.8	279	12	US-10-282-122A-45468	Sequence 45468, A
36	516	29.8	280	10	US-09-769-787-49	Sequence 49, Appli
37	516	29.8	280	12	US-10-282-122A-74062	Sequence 74062, A
38	516	29.8	317	12	US-10-425-114-66978	Sequence 66978, A
39	513	29.6	329	12	US-10-425-114-54493	Sequence 54493, A
40	513	29.6	379	12	US-10-425-114-59318	Sequence 59318, A
41	512	29.6	280	9	US-09-815-242-13290	Sequence 13290, A
42	511.5	29.5	368	16	US-10-038-854-127	Sequence 127, App
43	509.5	29.4	309	10	US-09-882-691-14	Sequence 14, Appli
44	509	29.4	298	15	US-10-369-493-13144	Sequence 13144, A
45	508.5	29.4	281	15	US-10-369-493-18317	Sequence 18317, A

ALIGNMENTS

RESULT 1
US-10-004-115A-1
; Sequence 1, Application US/10004115A
; Publication No. US20030134402A1
; GENERAL INFORMATION:
; APPLICANT: ASAKO, HIROYUKI
; APPLICANT: MATSUMURA, KENJI
; APPLICANT: SHIMIZU, NASATOSHI
; APPLICANT: ITO, NOBUYA
; APPLICANT: WAKITA, RUTHEI
; TITLE OF INVENTION: PROCESS FOR PRODUCING OPTICALLY ACTIVE
; FILE REFERENCE: 7372-72249
; CURRENT APPLICATION NUMBER: US/10/004,115A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: JP 2000-372704
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: JP 2001-006144
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: JP 2001-026594
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: JP 2001-175175
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Penicillium citrinum
US-10-004-115A-1

Query Match 100.0%; Score 1731; DB 14; Length 325;
Best Local Similarity 100.0%; Pred. No. 4.5e-170;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNGKFTLLSNGVKIPGVGF...FASGSGKGETYATVTTALKTGVRHLDCAWYLNNEGVG 60

Db 1 MSNGKFTLLSNGVKIPGVGF...FASGSGKGETYATVTTALKTGVRHLDCAWYLNNEGVG 60

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Onygenales; Microsporid Onygenales; Coccidioides.

OX NCBI_TaxID=5501;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Delgado N., Cole G.T.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AY059410; AAL27089.1; -

DR InterPro; IPR001395; Aldo/ket red.
 DR Pfam; PF00248; aldo ket red; 1.

DR PRINTS; PR00069; ALDKETREDTASE
 DR PROSITE; PD000288; Aldo/ket red; 1.

DR PROSITE; PS00798; ALDOKETO REDUCTASE 1; 1.
 DR PROSITE; PS00062; ALDOKETO REDUCTASE 2; 1.

SQ SEQUENCE 314 AA; 3541 MW; 09CD9B7734C34134 CRC64;

Query Match 30.8%; Score 533; DB 3; Length 314;
 Best Local Similarity 37.5%; Pred. No. 2.4e-35;

Matches 121; Conservative 61; Mismatches 109; Indels 32; Gaps 8;

QY 2 SNGKFTLSNGVPIGVGFTTFSESGKGTYYAVTTALTGTGVRHLDCAWYINVEGEVGE 61
 DB 3 SRHTFRLNTGCDIPALGFTGQWQDENAQED---AVLTALSAGYRHIDTAAYGTEKAIGR 59

QY 62 GIRDFLEKNSVKREDIFVCTKWNHLYRVEDVLWSTDDSLKRLGLDYDMLVHWPIA- 120
 DB 60 AL-----ARSLRDELFITSLKNNKHREDVEQALDQSLKLEIDYLDLKLHWPAAF 114

QY 121 -----AEKRGQGPFGKIPGPGKVIKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTTA 174
 DB 115 APGDDMFPKDSQGNKST-VDIDYV-----DTYKAMEKLKVGSKTKAIGISNFSK 163

QY 175 DLEKXSFPAKUMPHANOIEHPPLNBEIYQCFKSNIMPVAVSPIGSQNOVTTGCE--- 231
 DB 164 EMERLLESCISVPVAVMGMELHPLMQDNDVFWELKSGIHVTQYSSLGNOVEVSGRERYG 223

QY 232 RVSEKNTLNEIAEKGNTLAQVIAWGLRGVYVLPKSSNPKRIEBSNFKS-IELSDADFE 290
 DB 224 RLIEDPALAALGTYKGTGAQVALANGITKSHSVLVKSKTPHRIQQNFESDFELKQEDMK 283

QY 291 AINAVAKGRHFRFVNMDKDTGYD 313
 DB 284 AIENIDK--KLRFNDSSADFGYN 304

RESULT 10
 ID Q9VTK9 PRELIMINARY; PRT; 316 AA.

AC Q9VTK9;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE CG5084 protein (ID06393p).
 GN CG5084.

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;

RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Calle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Celniker S.,

Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,

RA Glöckle A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Moutkoulou G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J., Yao Q.A.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).

RL [2]
 RP SEQUENCE FROM N.A.

RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzales M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang X., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,

RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

RT "Annotation of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RL [4]
 RP SEQUENCE FROM N.A.

RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RL [5]
 RP SEQUENCE FROM N.A.

RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

RL [6]
 RP SEQUENCE FROM N.A.

RA STAPLETON M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,

RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,

RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;

QY 7 FTLSNGVKIPGVGFGTFASGSGKGETYAVTTTALKTGVRHLDCAWYVYLNAGEVGEGRDF 66
DB 15 FTLENGKXISIGLGTWRS--GKDETNAVCAALKAGYRHIDTAHIYGNKEKEIGEGIRE- 71
QY 67 LKENPSVKREDIFVCTKVNHLHRYEDVLSIDSLKRLGLDYDMLVHPWPIA----- 120
DB 72 ----SGVPRTDIWTSLKWCNAHRAGLVPLALEKTLQDLNLEYTDAYLIHWPALLSGPE 127
QY 121 -AEKNGGQEPKIPGDGVVILKDLTENPEPTWAMEKIYEDRKARSIGVSNWTIADLEKM 179
DB 128 ELPRNEKE-----LYEDVP--IETWQWELLTGTQVYIGISFNNEVILDRV 176
QY 180 SKFAYMHPANQIBIHPFLPNEELVQVCFKSNIMPVAYSPLGSONQVPTTG-BRVSENKT 238
DB 177 LKIAKVXPTLHQMELHPYLPQTEYLEKHKKLQIHVSAYSPLANQNDAYNSDISKLEHKT 236
QY 239 LNETAEKGGN--TLAOLVIAWGLRGGVVLPKSNPKRISNPKSIELSDADFEAINAVA 296
DB 237 LVDIANARGEGITPANTAISNAVAKRGTSVLPKSVNESRIVSNFLYIPLTDKMEAINNIG 296
QY 297 KGRHF-----RFVNMKD 308
DB 297 VVRFSGKPAKPMFVGLQD 317

RESULT 7

Q7ZA52 ID Q7ZA52 PRELIMINARY; PRT; 312 AA.
AC Q7ZA52;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Aldose reductase.
OS Magnaporthe grisea (Rice blast fungus) (Pyricularia grisea).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
OX NCBI_TaxID=148305;
RN [1]
RP SEQUENCE FROM N.A.
RA Zheng Y.-J., Tao Y., Zhang W., Jordan D.;
RT "Inhibition of fungal aldose reductase."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY033888; AAK55762.1; -;
SQ SEQUENCE 312 AA; 34368 MW; 10FE14BC676B5C55 CRC64;

Query Match 31.4%; Score 543; DB 3; Length 312;
Best Local Similarity 42.0%; Pred. No. 3.6e-36;
Matches 137; Conservative 53; Mismatches 110; Indels 26; Gaps 12;
QY 5 KFTLSNGVKIPGVGFGTFASGSGKGETYAVTTTALKTGVRHLDCAWYVYLNAGEVGEGR 64
DB 4 KDFLTNGAKIPAFGLGTW--QGDKGVKEAVLTAIKSGYRLIDGAVYVGNEEVGGIR 61
QY 65 DFLKENSVKREDIFVCTKW-NHLHRYEDVLSIDSLKRLGLDYDMLVHPWPIAAX 123
DB 62 EAISSG-IVKREDLVVSKWATYTTTCE---LGLDQSLKLLGLDYDMLVHPILMNP 117
QY 124 NQGE--PKIGPDGVVILKDLTENPEPTWAMEKIYEDRKARSIGVSNWTIADLEKMSK 181
DB 118 EGNDEKPK-KHADSRIIHH--THNHVDTWKLEKLPATGKTRAVGVSNYSKAWLQQLP 174
QY 182 FAKYNHPANQIBIHPFLPNEELVQVCFKSNIMPVAYSPLGSONQVPTTGERVSENKTNE 241
DB 175 HATTVPVAVNQHNPQLPQQLVDFCKEKGHIHAYVSPGSG-----TGGPILLTAEPVVK 228
QY 242 IAEKGGNTLAOLVIAWGLRGGVVLPKSNPKRISNPKSIELSDADFEAIN---AVAK 297
DB 229 IAEKHSISPAVLGVQIARITVIPSVPNPDRIKANAQLKDLDAEDMKLNDYSQAL 288
QY 298 -GRHFRFVNMKDTFGYDV-WPEETAK 321
DB 289 DGKLNRYV--PPRFGTDFGPDVKVG 312

RESULT 8

Q8LBG6 ID Q8LBG6 PRELIMINARY; PRT; 309 AA.
AC Q8LBG6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative NADPH dependent mannose 6-phosphate reductase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OX eursids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volkovskiy N., Town C.D., Troughan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation.";
RL Genome Biol. 0:0-0 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Brover V., Troughan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY087223; AAM64779.1; -;
DR InterPro; IPR001395; Aldo/ket red.
DR Pfam; PF00248; aldo ket red; I.
DR PRINTS; PR00069; ALDKETEDTASE.
DR ProDom; PD000288; Aldo/ket red; 1.
DR PROSITE; PS00798; ALDO-KETO-REDUCTASE_1; 1.
DR PROSITE; PS00062; ALDO-KETO-REDUCTASE_2; 1.
DR PROSITE; PS00063; ALDO-KETO-REDUCTASE_3; 1.
SQ SEQUENCE 309 AA; 34962 MW; 1FB3017045D86015 CRC64;
Query Match 30.8%; Score 533.5; DB 10; Length 309;
Best Local Similarity 40.1%; Pred. No. 2.1e-35;
Matches 120; Conservative 54; Mismatches 11; Indels 13; Gaps 7;
QY 8 TILSNGVKIPGVGFGTFASGSGKGETYAVTTTALKTGVRHLDCAWYVYLNAGEVGEGRDF 67
DB 4 TILSNGFKMIPVGLGVWRME--KEGIRDILNAIKIGYRHLDCAADYRNETEVDALTEAF 61
QY 68 KENPSVKREDIFVCTKVNHLHRYEDVLSIDSLKRLGLDYDMLVHPWPIAAXKNGQ 127
DB 62 KTG-LVKREDLFTTKLWNSDHG--VIEACKOSLUKLQDLDFLVHFPVATKHTGV 118
QY 128 --EPKIGPDGVVILKDLTENPEPTWAMEKIYEDRKARSIGVSNWTIADLEKMSKFAKV 185
DB 119 TTDSALGDDG--VLDDITTSILETTWHDWGLVSGYISNYDYFLTRDCLAYSKI 176
QY 186 MHPANQIBIHPFLPNEELVQVCFKSNIMPVAYSPLGSONQVPTTGERVS--ENKTLNIA 243
DB 177 KEAVNQIETHPTFYQDRLVFKQKHGICVTAHTPLGGATANAEMFGTVCSDDDPVLKDDA 236
QY 244 EKGGGNTLAOLVIAWGLRGGVVLPKSNPKRISNPKSIELSDADFEAINAVAKGRH 300
DB 237 EYKKEVTAQVLRWGIQKTVIPKTPARLEENFQVDFELSKEDNEVSKMERKYH 295

RESULT 9

Q96UH3 ID Q96UH3 PRELIMINARY; PRT; 314 AA.
AC Q96UH3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Aldohyde reductase.
GN ALDI.
OS Coccidioides immitis.

RL Yeast 0:0-0(2001).
DR EMBL; AB047396; BABL1959.1; --.
DR HSP; P06632; 1HW6.
DR InterPro; IPR001395; Aldo/ket red.
DR Pfam; PF00248; aldo ket red; 1.
DR PRINTS; PR00069; ALDKETREDTASE.
DR ProDom; PD000288; Aldo/ket red; 1.
DR PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
DR PROSITE; PS00062; ALDOKETO_REDUCTASE_2; 1.
DR PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.
SQ SEQUENCE 310 AA; 34126 MW; F201FFBF8CCB9CFD CRC64;

Query Match 32.1%; Score 555; DB 3; Length 310;
Best Local Similarity 40.5%; Pred. No. 3.7e-37;
Matches 126; Conservative 48; Mismatches 93; Indels 44; Gaps 9;

QY 6 TFLNSGVKIPGVGFGTFASEGSGEYTYTAVTALKTGYHLOCAYWYLNNEGVGEGIRD 65
DB 10 TLKLTGQTTPQVGLGTRSGENEG--YKAVIEALKAGYRHIDCAAYVNGEVGKAIQD 67
QY 66 FLKENPSVKREDIFVCTKVW-NHLHRYEDVLWSDSLKELGLDYDMFLVHPIA---- 120
DB 68 -----SGVPRNEIFLTKWCTHQRNQEAL----DQSLQLGLDYGLVLYVHPVPERTE 119
QY 121 -----AEKNGQGEPIKGPDGKYVILKLTENPEPTWRAMEKIYEDRKARSIGVSN 170
DB 120 NIKDGNLFQFAEK-----PDGSKDV--DLEWNFIKTWELMQKLESCKTKAIGVSN 168
QY 171 WTADLEKM--SKFAKMPHANQIEHPFLPNEELVOYCFSKNIMPVAYSPILGSONQVPT 228
DB 169 FSVNNKDLALAAPTTKVTPAVNVEFHLLPQBELIQYCEKGVIEAYSPILGSEN---- 224
QY 229 TGERVSENKTLNEIAEKGNTLAQVLIAMGLRGRGVVLPKSSNPKRIESNFKSIELSDAD 288
DB 225 --APILSDPTVOELAKNGVAGHVVISWAVQKGLVTLPSKSVTPSRIVGNLKVILSDSD 282
QY 289 FEAINAVAKGR 299
DB 283 VAKVDGLLKK 293

RESULT 5
P74308
ID P74308 PRELIMINARY; PRT; 327 AA.
AC P74308;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DE Aldehyde reductase.
GN SLR0942.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asanizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90914; BAA18402.1; --.
DR PIR; S76143; S76143.
DR HSP; P14550; 2ALR.
DR InterPro; IPR001395; Aldo/ket red.
DR Pfam; PF00248; aldo ket red; 1.
DR PRINTS; PR00069; ALDKETREDTASE.
DR ProDom; PD000288; Aldo/ket red; 1.
DR PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.

DR PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.
SQ SEQUENCE 327 AA; 36014 MW; 4B9415E098A8892D CRC64;

Query Match 31.9%; Score 553; DB 16; Length 327;
Best Local Similarity 38.1%; Pred. No. 5.9e-37;
Matches 117; Conservative 69; Mismatches 105; Indels 16; Gaps 6;

QY 5 KTFILNSGVKIPGVGFGTFASEGSGEYTYTAVTALKTGYHLOCAYWYLNNEGVGEGIR 64
DB 11 KYFPLNSGEQIPALGLGTWKS--SPQVVQQAIVEQALDGYRHLDCAAIYNEAIGATLA 68
QY 65 DFLKENPSVKREDIFVCTKVW-NHLHRYEDVLWSDSLKRLGLDYDMFLVHPIAAEKN 124
DB 69 NAFTKG-VVKEELWITSKLSNAHPDAVLPALEKTLQDLGLDYDLVLIHWPVVI--- 124
QY 125 GQGEPIKG-PDGKYVILKLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMSKFA 183
DB 125 ---QPDVGFPESGDQLLPFTFASLEGTMQALEKAVDLGLCHHIGVSNFSLKLEMLVLSMA 181
QY 184 KMPHANQIEHPFLPNEELVOYCFSKNIMPVAYSPILGSONQVPTTGERVSENKTL---- 239
DB 182 RIPPAVQVELHPYLOQSDLLTFANSQNILLTAYSPILGSGDR-PAATQQAEPKLLTDPV 240
QY 240 -NETAEKCGNTLAQVLIAMGLRGRGVVLPKSSNPKRIESNFKSIELSDADFEAINAVAKG 298
DB 241 INGIAAEQCSAAQVLLAWAIQRTGTTPKSNVPERLEQNLRADITLTDSEMAKIALLD 300
QY 299 RHEFEVN 305
DB 301 RHYRIVS 307

RESULT 6
O42888
ID O42888 PRELIMINARY; PRT; 325 AA.
AC O42888;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Probable oxidoreductase.
GN SPBC8E4.04.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Connor R., Churche C.M.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021815; CAA16997.1; --.
DR PIR; T50378; T39169.
DR HSP; P14550; 2ALR.
DR GeneDB Spombe; SPBC8E4.04; --.
DR InterPro; IPR001395; Aldo/ket red.
DR Pfam; PF00248; aldo ket red; 1.
DR PRINTS; PR00069; ALDKETREDTASE.
DR ProDom; PD000288; Aldo/ket red; 1.
DR PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
DR PROSITE; PS00062; ALDOKETO_REDUCTASE_2; 1.
DR PROSITE; PS00063; ALDOKETO_REDUCTASE_3; FALSE NEG.
SQ SEQUENCE 325 AA; 36629 MW; 68ED211F1E4D08D3 CRC64;

Query Match 31.7%; Score 548.5; DB 3; Length 325;
Best Local Similarity 40.5%; Pred. No. 1.4e-36;
Matches 130; Conservative 52; Mismatches 102; Indels 37; Gaps 8;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 14:00:41 ; Search time 45 Seconds
(without alignments)
2278.743 Million cell updates/sec

Title: US-10-004-115B-1

Perfect score: 1731

Sequence: 1 MSGNGKFTLSNGVKIPGVGF.....MKDTFGYVWPETAKNLSA 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1514	87.5	325	3 Q877A2	Q877A2 aspergillus
2	1471	85.0	325	3 Q728L1	Q728L1 emericeila
3	1037	59.9	254	3 Q74646	Q74646 gibberella
4	555	32.1	310	3 Q9HG10	Q9HG10 zygosacchar
5	553	31.9	327	16 P74308	P74308 synecocyst
6	548.5	31.7	325	3 Q42888	Q42888 schizosacch
7	543	31.4	312	3 Q7ZA52	Q7ZA52 magnaporthe
8	533.5	30.8	309	10 Q8LHG6	Q8LHG6 arabidopsis
9	533	30.8	314	3 Q6UJH3	Q6UJH3 coccidioid
10	531.5	30.7	316	5 Q9VTK9	Q9VTK9 drosophila
11	530.5	30.6	279	16 Q8LB04	Q8LB04 bacillus ce
12	530	30.6	310	3 Q9HGK9	Q9HGK9 zygosacchar
13	530	30.6	322	10 Q8GXW0	Q8GXW0 arabidopsis
14	529.5	30.6	309	10 Q9SJV2	Q9SJV2 arabidopsis
15	529	30.6	290	10 Q80945	Q80945 arabidopsis
16	528	30.5	280	16 Q8E460	Q8E460 streptococc

17	528	30.5	280	16 Q8DYK4	Q8DYK4 streptococc
18	527.5	30.5	309	10 Q9SJV1	Q9SJV1 arabidopsis
19	526	30.4	315	10 Q9AW93	Q9AW93 digitalis p
20	525.5	30.4	350	5 Q8IQF8	Q8IQF8 drosophila
21	525	30.3	279	16 Q81XD1	Q81XD1 bacillus an
22	525	30.3	279	16 Q81S55	Q81S55 bacillus ce
23	525	30.3	315	10 Q9AW92	Q9AW92 digitalis p
24	523.5	30.2	313	10 Q82020	Q82020 medicago sa
25	523	30.2	286	16 Q88SL1	Q88SL1 lactobacill
26	522.5	30.2	275	16 Q818Y9	Q818Y9 bacillus ce
27	522.5	30.2	311	10 Q80944	Q80944 arabidopsis
28	521	30.1	276	5 Q9GV41	Q9GV41 trypanosoma
29	521	30.1	320	10 Q9M009	Q9M009 arabidopsis
30	519	30.0	280	16 Q8DUR6	Q8DUR6 streptococc
31	518.5	30.0	275	16 Q81MD0	Q81MD0 bacillus an
32	518.5	30.0	309	10 Q8LD92	Q8LD92 arabidopsis
33	518	29.9	274	16 Q8EST2	Q8EST2 oceanobacil
34	516.5	29.8	320	5 Q9VZK8	Q9VZK8 drosophila
35	516	29.8	279	16 Q81MX4	Q81MX4 bacillus an
36	516	29.8	280	16 Q97PW2	Q97PW2 streptococc
37	516	29.8	280	16 Q8DF64	Q8DF64 streptococc
38	515.5	29.8	304	10 Q84W94	Q84W94 arabidopsis
39	508.5	29.4	281	16 Q9CIT8	Q9CIT8 lactococcus
40	507	29.3	276	5 Q8IUI5	Q8IUI5 crithidia f
41	506.5	29.3	274	5 Q965C7	Q965C7 trypanosoma
42	503	29.1	276	16 Q32210	Q32210 bacillus su
43	503	29.1	310	10 Q9FVN7	Q9FVN7 orobanche r
44	500	28.9	280	16 Q34678	Q34678 bacillus su
45	498	28.8	274	16 Q836H3	Q836H3 enterococcu

ALIGNMENTS

RESULT 1

Q877A2
ID Q877A2 PRELIMINARY; PRT; 325 AA.
AC Q877A2;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Reductase-like protein.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI840;
RA Akao T., Akeno T., Goto K., Akita O.,
RT "cloning and nucleotide sequencing of cDNA that encodes a reductase-
like protein from Aspergillus oryzae."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB079030; SACS6099.1;
DR InterPro; IPR001395; Aldo/ket_red.
DR Pfam; PF00248; aldo ket red; 1.
DR PRINTS; PR00069; ALDKETREDTASE.
DR PRODOM; PD000288; Aldo/ket red; 1.
DR PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
SQ SEQUENCE 325 AA; 36932 MW; DA5ED3898D0C4BD9 CRC64;

Query Match 87.5%; Score 1514; DB 3; Length 325;
Best Local Similarity 85.8%; Pred. No. 1.2e-115;
Matches 279; Conservative 22; Mismatches 24; Indels 0; Gaps 0;

QY	1	MSGNGKFTLSNGVKIPGVGF	FTASEGSGKGTGTAVTT	TALKTGYRHLDCAWYLN	GEVSG 60
Db	1	MSGNGKFTLSNGVKIPGVGF	FTASEGSGKGTGTAVTK	QAVTKALETGYRHLDCAW	FLNDEVG 60
QY	61	EGIRDFLKENSVKREDIF	VCCTKVNHLHRYEDV	VLMSIDSLKRLGLDYDM	FLVHPPIA 120
Db	61	DGIHDFLKFNFSVKREDI	FVCTKVNHLHRLPED	VQWSVDNSLKRRLRLDY	VDLFLVHPPIA 120

QY 291 AINAVAK-GRHFRFVNVMKOTFGYDVWP 311
DB 295 EI---AKLDRHLRF---NDPFWTKIP 315

Search completed: May 28, 2004, 14:04:47
Job time : 19 secs

XY11 PACTA
ID XYL1_PACTA STANDARD; PRT; 318 AA.
AC P78736;
DT 01-NOV-1997 (Rel. 35; Created)
DT 01-NOV-1997 (Rel. 35; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE NAD(P)H-dependent xylose reductase [EC 1.1.1.-] (XR).
OS Pachysolen tannophilus (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pachysolen.
NCBI_TaxID=4918;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=NRRL Y-2460 / ATCC 32691;
RC MEDLINE=97082506; PubMed=8923742;
RX Bolen P.L., Hayman G., Shepherd H.S.;
RA "Sequence and analysis of an aldose (xylose) reductase gene from the
RT xylose-fermenting yeast Pachysolen tannophilus.";
RL Yeast 12:1367-1375(1996).
RN [2]
RN SEQUENCE OF 1-62.
RP Bolen P.L., Bietz J.A., Detroy R.W.;
RA "Aldose reductase in the yeast Pachysolen tannophilus: purification,
RT characterization and N-terminal sequence."
RT Bio-technol. Bioeng. Symp. 15:129-148(1985).
CC 1- FUNCTION: Reduces xylose into xylitol.
CC 1- PATHWAY: D-xylose degradation.
CC 1- SIMILARITY: Belongs to the aldo/keto reductase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U40706; AAC49526.1; -
CC HSP; P14550; 2ALR.
DR InterPro: IPR001395; Aldo/ket red.
DR Pfam: PF00248; aldo_ket_red; 1.
DR PRINTS; PRO0069; ALDKETRDIASE.
DR ProDom; PD000288; Aldo/ket_red; 1.
DR PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
DR PROSITE; PS00062; ALDOKETO_REDUCTASE_2; FALSE NEG.
DR PROSITE; PS00063; ALDOKETO_REDUCTASE_3; FALSE_NEG.
KW Oxidoreductase; Xylose metabolism; NAD.
FT ACT_SITE 48 48 HYDROGEN-BOND DONOR (BY SIMILARITY).
FT SEQUENCE 318 AA; 36422 MW; 891FC6F461D231EC CRC64;

Query Match 28.1%; Score 486; DB 1; Length 318;
Best Local Similarity 38.2%; Pred. No. 6.2e-31;
Matches 125; Conservative 51; Mismatches 117; Indels 34; Gaps 11;
QY 7 FTLSNGVPIPGVCGFTPASESGSKSETYTAVTTALKTGYRHLDCAWYVINEGEVGEGRDF 66
DB 6 YTLNNGRKIPATPGMGCWKLNAADVMVAAI---KEGYRLFDCACDYGNKEKEVGEGINFA 61
QY 67 LKENPSVKREDIFVCTKVNHLHRYEDVLMISDLSLRLGLDYVDMFLVHWPIAAEKNGQ 126
DB 62 IKDG-LVKKRLFTITSLNNFHAKENVKKALMSLSDFNLDYFDLYMHFPISF-KFVP 119
QY 127 GEPKIGP-----DGKVIWKDLTENPEP---TWAMEKIYEDRKARSIGVSNWTIADLEK 178
DB 120 FEETPPPGFYCGDGFYEDV-----PIETWRAMENLVDEGLVKISGVSNVSGGLED 174
QY 179 MSKFAWPHANQIIEHPFLPNEELVQYCFSKNTMPVAYSPLGS-----QNOVPTTGR 232
DB 175 LTKARIRPASQIEHHFYLOQNLKIVEAQLKGIWVTGYTSNFGPLSLFLGNETAKTQP 234
QY 233 VSENKTLNEIAEKGNTLAQVLIAGWLRGRGVVLPKSNPKRIESNF--KSIELSDADFE 290
DB 235 LYENKTIITIAAKHGKTFQVLLRWVNGRGTATIPKSTFTFNTLAVNLHVDFDLTKEDFE 294

QY 291 AINAVAK-GRHFRFVNVMKOTFGYDVWP 311
DB 295 E1---AKLDRHLRF---NDPFWTKIP 315

Search completed: May 28, 2004, 14:04:47
Job time : 19 secs

XY11 PACTA
ID XYL1_PACTA STANDARD; PRT; 318 AA.
AC P78736;
DT 01-NOV-1997 (Rel. 35; Created)
DT 01-NOV-1997 (Rel. 35; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE NAD(P)H-dependent xylose reductase [EC 1.1.1.-] (XR).
DE Pachysolen tannophilus (Yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pachysolen.
NCBI_TaxID=4918;
OX X
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=NRRL Y-2460 / ATCC 32691;
RC MEDLINE=97082506; PubMed=8923742;
RX Bolen P.L., Hayman G., Shepherd H.S.;
RA "Sequence and analysis of an aldose (xylose) reductase gene from the
RT xylose-fermenting yeast Pachysolen tannophilus.";
RL Yeast 12:1367-1375(1996).
RN [2]
RN SEQUENCE OF 1-62.
RP Bolen P.L., Bietz J.A., Detroy R.W.;
RA "Aldose reductase in the yeast Pachysolen tannophilus: purification,
RT characterization and N-terminal sequence."
RT Bio-technol. Bioeng. Symp. 15:129-148(1985).
CC 1-! FUNCTION: Reduces xylose into xylitol.
CC 1-! PATHWAY: D-xylose degradation.
CC 1-! SIMILARITY: Belongs to the aldo/keto reductase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U40706; AAC49526.1; -
CC HSP; P14550; 2ALR.
DR InterPro: IPR001395; Aldo/ket red.
DR Pfam: PF00248; aldo_ket_red; 1.
DR PRINTS; PRO0069; ALDKETRDIASE.
DR ProDom; PD000288; Aldo/ket_red; 1.
DR PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
DR PROSITE; PS00062; ALDOKETO_REDUCTASE_2; FALSE NEG.
DR PROSITE; PS00063; ALDOKETO_REDUCTASE_3; FALSE_NEG.
KW Oxidoreductase; Xylose metabolism; NAD.
FT ACT_SITE 48 48 HYDROGEN-BOND DONOR (BY SIMILARITY).
FT SEQUENCE 318 AA; 36422 MW; 891FC6F461D231EC CRC64;

Query Match 28.1%; Score 486; DB 1; Length 318;
Best Local Similarity 38.2%; Pred. No. 6.2e-31;
Matches 125; Conservative 51; Mismatches 117; Indels 34; Gaps 11;
QY 7 FTLSNGVPIPGVCGFTFASGSGKSETTYTAVTTALKTGYRHLDCAWYVINEGEVGEGRDF 66
DB 6 YTLNNGRKIPATIGMGCWKLNAADVMVAAI---KEGYRLFDCACDYGNKEKEVGEGINFA 61
QY 67 LKENPSVKREDIFVCTKVNHLHRYEDVLMISDLSLRLGLDYVDMFLVHWPIAAEKNGQ 126
DB 62 IKDG-LVKKRLFTISKLNWNNFHAKENVKKALMSLSDFNLDYFDLYLHFFISF-KFVP 119
QY 127 GEPKIGP-----DGKVVILKDLTENPEP---TWAMEKIYEDRKARSIGVSNWTIADLEK 178
DB 120 FEETPPPGFYCGDGGKFYEDV-----PIETWRAMENLVDEGLVKSIGVSNVSGGLED 174
QY 179 MSKFAWPHANQIIEHPFLPNEELVQYCFSKNTMPVAYSPLGS-----QNOVPTTGR 232
DB 175 LTKARIRPASQIEHHFYLOQKLVKVEAQLKGIWVTGYTSNFGPLSLFLGNETAKTQP 234
QY 233 VSENKTLNEIAEKGNTLAQVLIWGLRGRGVVLPKSNPKRIESNF--KSIELSDADFE 290
DB 235 LYENKTTITIAAKHGKTFQVLLRWVNGRGTATIPKSTFTFNTLAVNLHVDFDLTKEDFE 294

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[9] SEQUENCE OF 243-274.
RA MEDLINE=89123393; PubMed=2492527;
RX Morjana N.A., Lyons C., Flynn T.G.;
RA "Aldehyde reductase from human psoas muscle. Affinity labeling of an
RT active site lysine by pyridoxal 5'-phosphate and pyridoxal
RT 5'-diphospho-5'-adenosine.";
RL J. Biol. Chem. 264:2912-2919(1989).
[10]
RA PARTIAL SEQUENCE, AND ACETYLATION.
RC TISSUE-Muscle.
RX MEDLINE=94109388; PubMed=8281941;
RA Aquilino M., Potier N., Klarskov K., Reymann J.-M., Sorokine O.,
RA Kieffer S., Barth P., Andrianomanga V., Biellmann J.-P.,
RA van Dorselaer A.;
RT "Sequence of pig lens aldose reductase and electrospray mass
RT spectrometry of non-covalent and covalent complexes.";
RL Eur. J. Biochem. 218:893-903(1993).
[11]
RA MUTAGENESIS OF ASP-43; TYR-48; LYS-77 AND HIS-110.
RX MEDLINE=94064641; PubMed=8245005;
RA Tarle I., Borhani D.W., Wilson D.K., Quirocho F.A., Petrash J.M.;
RT "Probing the active site of human aldose reductase. Site-directed
RT mutagenesis of Asp-43, Tyr-48, Lys-77, and His-110.";
RL J. Biol. Chem. 268:25687-25693(1993).
[12]
RA X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
RX MEDLINE=92320300; PubMed=1621098;
RA Wilson D.K., Bohren K.M., Gabbay K.H., Quirocho F.A.;
RT "An unlikely sugar substrate site in the 1.65 Å structure of the
RT human aldose reductase holoenzyme implicated in diabetic
RT complications.";
RL Science 257:81-84(1992).
[13]
RA X-RAY CRYSTALLOGRAPHY (2.27 ANGSTROMS).
RX MEDLINE=93077587; PubMed=1447221;
RA Borhani D.W., Harter T.M., Petrash J.M.;
RT "The crystal structure of the aldose reductase.NADPH binary complex.";
RL J. Biol. Chem. 267:24841-24847(1992).
[14]
RA X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=94052189; PubMed=8234324;
RA Wilson D.K., Tarle I., Petrash J.M., Quirocho F.A.;
RT "Refined 1.8-Å structure of human aldose reductase complexed with the
RT potent inhibitor zopolrestat.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:9847-9851(1993).
[15]
RA X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=98070237; PubMed=9405046;
RA Harrison D.H., Bohren K.M., Petsko G.A., Ringe D., Gabbay K.H.;
RT "The alarstatin double-decker: binding of two inhibitor molecules to
RT human aldose reductase reveals a new specificity determinant.";
RL Biochemistry 36:16134-16140(1997).
CC -!- FUNCTION: Catalyzes the NADPH-dependent reduction of a wide
CC variety of carbonyl-containing compounds to their corresponding
CC alcohols with a broad range of catalytic efficiencies.
CC -!- CATALYTIC ACTIVITY: Alditol + NAD(P)(+) = aldose + NAD(P)H.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- DISEASE: In diabetes and galactosemia, increased AR activity leads
CC to high levels of sorbitol and galactitol, respectively, in the
CC cells of many tissues. Accumulation of sugar alcohols has been
CC shown to cause osmotic catarracts in the lens. AR is also thought
CC to play a key role in diabetic complications of three other target
CC tissues, namely, nerve, kidney and retina.
CC -!- SIMILARITY: Belongs to the aldo/keto reductase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X15414; CAA33460.1; -
CC EMBL: J04795; AAA51713.1; -
CC EMBL: J05017; AAA51714.1; -
CC EMBL: M34720; AAA35560.1; -
CC EMBL: M34721; AAA35561.1; -
CC EMBL: J05474; AAA51715.1; -
CC EMBL: M59783; AAA51712.1; -
CC EMBL: M59856; AAA51712.1; JOINED.
CC EMBL: AF032455; AAB88851.1; -
CC EMBL: BC000260; AAH00260.1; -
CC EMBL: BC005387; AAH05387.1; -
CC EMBL: BC010391; AAH10391.1; -
CC PIR: A39763; A39763.
CC PDB: 1ADS; 31-OCT-93.
CC PDB: 1ABN; 31-JAN-94.
CC PDB: 2ACQ; 31-JUL-94.
CC PDB: 2ACR; 31-JUL-94.
CC PDB: 2ACS; 31-JUL-94.
CC PDB: 2ACU; 31-JUL-94.
CC PDB: 1MAR; 20-JUL-95.
CC PDB: 1AZ1; 18-MAR-98.
CC PDB: 1AZ2; 18-MAR-98.
CC PDB: 1ZP3; 07-FEB-01.
CC PDB: 1EL3; 10-MAY-00.
CC Aarhus/Ghent-2DPAGE; 1202; IEF.
CC Genew; HGNC:381; AKR1B1.
CC MIM: 103880; -
CC GO: GO:0005615; C:extracellular space; TAS.
CC GO: GO:0004032; F:aldehyde reductase activity; TAS.
CC GO: GO:0005483; F:electron transporter activity; TAS.
CC GO: GO:0005975; F:carbohydrate metabolism; TAS.
CC InterPro: IPR001395; Aldo/ket red.
CC Pfam: PF00248; Aldo/ket red; 1.
CC PRINTS: PR00069; ALDKETREDTASE.
CC PRODOM: PD000288; Aldo/ket red; 1.
CC PROSITE: PS00062; ALDO-KETO-REDUCTASE_2; 1.
CC PROSITE: PS00063; ALDO-KETO-REDUCTASE_3; 1.
CC PROSITE: PS00798; ALDO-KETO-REDUCTASE_1; 1.
CC Oxidoreductase; NADP; Acetylation; 3D-structure; Polymorphism.

Query Match 28.1%; Score 486.5; DB 1; Length 315;
Best Local Similarity 38.1%; Pred. No. 5.6e-31;
Matches 117; Conservative 51; Mismatches 110; Indels 29; Gaps 8;

QY 9 LNSGVKIPGVGFTFASGSGKGTVTATTALKGYRHLDCAWYLYNVEGVEGIRDFLK 68
DB 6 LNSGAKPILGLGTWKS--PGQVTEAVKVAIDVGYRHIDCAHYVQNEVGVATQEKLR 63
QY 69 ENPSVKREDIFVCTKVMNHLRYEDVLWSIDSLKRLGLDYDMELVHWPAAEKNGQGE 128
DB 64 EQ-VVRELFVSKLWCTYHEKGLVKGCOKTISDLKLDLYLHLHWPTGFK----- 116
QY 129 PKIGPDGKYLKDLTENPEP-----TWAMEKIYEDRKARSIGVSNWTADLEKM-- 179
DB 117 -----PGKEFFELDESNNVPSDNTNILDITWAAEBELVDEGLVKAIGISNFHQLQVEMTLN 171
QY 180 SKFAKVMHANOIEIHPLELVOYCFNSNIMPVAVSLGSONQVPTTGERVY--ENK 237
DB 172 KFLGLYKPAVNOIECHPYLTQELQYQCSQGI VVTAYSLPGSPDPWAKEDPSLLSDP 231
QY 238 TLNEIAEKGNTLAQVLIAGLGRGYVLPKSSNPKRIESNFK--SIELSDADFEAINAV 295
DB 232 RIKATAAKHNKTTAQVLLRFPMQRLNVLVIPSVPTEAENFKVDFELSSQDMTTL--L 289
QY 296 AKGRHFR 302
DB 290 SYNRNWR 296

RESULT 15

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U37100; AAC17469.1; --
CC DR EMBL; AF052577; AAC36465.1; --
CC DR EMBL; BC008337; AAH08837.1; --
CC DR EMBL; AF044961; AAC15671.1; --
CC DR Genew; HGNC:382; AKR1B10.
CC MIM; 604707; --
CC DR HSP; P45377; 1FRB.
CC DR GO; GO:0004033; F:aldo-keto reductase activity; TAS.
CC DR InterPro; IPR001395; Aldo/ket red.
CC DR Pfam; PF00248; aldo ket red; 1.
CC DR PRINTS; PR00069; ALDOKETREDASE.
CC DR ProDom; PD000288; Aldo/ket red; 1.
CC DR PROSITE; PS00798; ALDOKETO-REDUCTASE 1; 1.
CC DR PROSITE; PS00062; ALDOKETO-REDUCTASE 2; 1.
CC DR PROSITE; PS00063; ALDOKETO-REDUCTASE 3; 1.
CC DR Oxidoreductase; NADP; Polymorphism.
CC FT ACT SITE 49 49 HYDROGEN-BOND DONOR (BY SIMILARITY).
CC FT VARIANT 313 313 D -> N.
CC FT SEQUENCE 316 AA; 36020 MW; 0C2FC0E378955A33 CRC64;
CC /FTID=VAR_013287.
CC
CC Query Match 28.1%; Score 487; DB 1; Length 316;
CC Best Local Similarity 36.6%; Pred. No. 5.2e-31;
CC Matches 118; Conservative 58; Mismatches 122; Indels 24; Gaps 8;
CC
Qy 6 TFT-LSNGKIPGVGFGTFSESGKGTGTAVTALTKVGRHLDCAWYVNEGEVGEGR 64
Db 3 TFVELSTKAKPIVGLTWKS--PLGKVKAEAKVADAGVRHIDCAVYQNEVEGEATQ 60
Qy 65 DFLKENPSVKREDIFVCTKVNNHLHYEDVLSIDSLKRLGLDYDMFLVHWPIAAKN 124
Db 61 EKIQEK-AVKREDLFVSKLWPTFFERPLVRKAPEKTLKDLKSLYDLVLIHWPFQKSG 119
Qy 125 GQGEPK-----IGPDQKYVLKDLTENPEPTVRAMEKIYEDRKASIGVSNWTIADLEK 178
Db 120 DDLFPDKDKNAIG--GKATFL-----DAWEAMEELVDEGLVAKGVSNFSHFQIEK 169
Qy 179 M--SKFAKWPHANQIHFPLNEELVQYCFKKNIMPVAYSPLGSONQVPTTGERVS-- 234
Db 170 LLNKPGLKYPVTVNQVECHPYLQEKLIQYCHSGKITVTAYSPLGSPDRPWAKPDPSSL 229
Qy 235 ENKTLNIEAKGGMNTLAQVLIANGRLRGYVVLPKSNPKRIESNFKSLSDADFAINA 294
Db 230 EDPKIEIAAKHKHTAAQVLIIRFHQIRNVIPKSVTPAIRIVENIQVDFPKLSDEEMATI 289
Qy 295 VAKGRHFRFVNMKDTFGYDVMP 316
Db 290 LSFNFRNWRACNVLOSSHLEDP 311
CC
RESULT 14
ALDR HUMAN
ID -ALDR HUMAN STANDARD; PRT; 315 AA.
AC P15121; O9BS21;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Aldehyde reductase (EC 1.1.1.21) (AR) (Aldehyde reductase).
GN AKR1B1 OR ALDR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89255461; PubMed=2498333;
RA Bohren K.M., Bullock B., Wermuth B., Gabbay K.H.;
RT "The aldo-keto reductase superfamily. cDNAs and deduced amino acid
RN sequences of human aldehyde and aldose reductases.";
RN J. Biol. Chem. 264:9547-9551(1989).
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal;
RX MEDLINE=90045960; PubMed=2510130;
RA Graham A., Hedge P.J., Powell S.J., Riley J., Brown L., Gammack A.,
RA Carey F., Markham A.F.;
RT "Nucleotide sequence of cDNA for human aldose reductase.";
RN Nucleic Acids Res. 17:8368-8368(1989).
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=89359274; PubMed=2504709;
RA Chung S., Lamendola J.;
RT "Cloning and sequence determination of human placental aldose
RN reductase Gene.";
RN J. Biol. Chem. 264:14775-14777(1989).
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=90253609; PubMed=2111143;
RA Grundmann U., Bohn H., Obermeier R., Amann E.;
RT "Cloning and prokaryotic expression of a biologically active human
RN placental aldose reductase.";
RN DNA Cell Biol. 9:149-157(1990).
RP SEQUENCE FROM N.A.
RX MEDLINE=90277668; PubMed=2112546;
RA Nishimura C., Matsura Y., Kokai Y., Akera T., Carper D., Morjana N.,
RA Lyons C., Flynn T.G.;
RT "Cloning and expression of human aldose reductase.";
RN J. Biol. Chem. 265:9788-9792(1990).
RP SEQUENCE FROM N.A.
RX MEDLINE=91201333; PubMed=1901857;
RA Graham A., Brown L., Hedge P.J., Gammack A.J., Markham A.F.;
RT "Structure of the human aldose reductase gene.";
RN J. Biol. Chem. 266:6872-6877(1991).
RP SEQUENCE FROM N.A.
RX MEDLINE=97341182; PubMed=9195951;
RA Ko B.C.B., Ruepp B., Bohren K.M., Gabbay K.H., Chung S.S.;
RT "Identification and characterization of multiple osmotic response
RN sequences in the human aldose reductase gene.";
RN J. Biol. Chem. 272:16431-16437(1997).
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Eye, and Urinary bladder;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Soufaratne P.H.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RN human and mouse cDNA sequences."

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to licenses@isb-sib.ch).

[illegible]

```

DR EMBL; U13691; AAB60687.1; JOINED.
DR EMBL; U13692; AAB60687.1; JOINED.
DR EMBL; U13693; AAB60687.1; JOINED.
DR EMBL; M32818; AAA31160.1; -.
DR EMBL; U12316; AAA50833.1; -.
DR EMBL; J05048; AAA31157.1; -.
DR PIR; A34406; A34406.
DR HSP; P15121; 2ACQ.
DR InterPro; IPR001395; Aldo/ket red.
DR Pfam; PF00248; aldo_ket_red; 1.
DR PRINTS; PR00069; ALDKETREDTASE.
DR ProDom; PD000288; Aldo/ket red; 1.
DR PROSITE; PS00062; ALDOKETO REDUCTASE 2; 1.
DR PROSITE; PS00063; ALDOKETO REDUCTASE 3; 1.
DR PROSITE; PS00798; ALDOKETO REDUCTASE_1; 1.
KW Oxidoreductase; NADP; Acetylation.
FT INIT MET 0 0 BY SIMILARITY.
FT NP BIND 9 18 NADP (POTENTIAL).
FT ACT SITE 48 48 HYDROGEN-BOND DONOR (BY SIMILARITY).
FT MOD RES 1 1 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 315 AA; 35632 MW; 64D53E6AC0853FEB CRC64;

Query Match 28.2%; Score 488.5; DB 1; Length 315;
Best Local Similarity 38.4%; Pred. No. 3.9e-31;
Matches 118; Conservative 49; Mismatches 111; Indels 29; Gaps 8;

QY 9 LNSGKIPGVGFGFASGSGKGTYYTAVTTALKTGYRHLDCAWYLYNEGEVGEIGRIDFLK 68
DB 6 LYGAKMPLILGLTWKS--PPGQVTEAVKTAIDLGYRHIDCAHVQNEVEGVALQEK 63

QY 69 ENSVSKREDIPVCKVWNHHRVEDVWLSIDDSKLGLDYDVMFLVHWPFAEKNGQGE 128
DB 64 EQ-VVKEELFVSKLWCTSHDKSLVKGACQKTLNDLKLDTLYLIHWPTGFGHSEYF 122

QY 129 PKIGPDCKYVILKDLTENPSP-----TWRAEKIVEDRKASIGVSNWTIADLEKMA-- 179
DB 123 P-----LDAGNVIPSTDFDTWEAGLVDGLVKSIGVSNFNLQIERILN 171

QY 180 SKFAKVMPHANQIEIHPFLPNEBLVQYCFKSNIMPVAYSLGSSQVPTTGERVS--ENK 237
DB 172 KPLGKYKPAVNVIECHPYLTOEKLQYCHSGKIWTVAYSLGSPDRFWAKPEDPSLLEDP 231

QY 238 TLNEIAKGGNTLAQVLIANGRGVYVLPKSNPKRIENFK--SIELSDADFEAINAV 295
DB 232 RIKAIADKHKKTAQVLIIRPQMORNLVDPKSVTPARIAENFQVFFELSSDDMTTL--L 289

QY 296 AKGRHFR 302
DB 290 SYNEVNR 296

RESULT 12
ID ALDR_MOUSE STANDARD; PRT; 315 AA.
AC P45376; O70130; Q99K9;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Aldo reductase (EC 1.1.1.21) (AR) (Aldehyde reductase).
GN AKR1B1 OR ALDR1 OR ALDR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Kidney;
RX MEDLINE=95154325; PubMed=7851421;
RA Gui T., Tanimoto T., Kokai Y., Nishimura C.;
RT "Presence of a closely related subgroup in the aldo-ketoreductase family of the mouse."
RL Eur. J. Biochem. 227:448-453 (1995).
[2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ICR X Swiss Webster; TISSUE=Liver;
RA Iwata T., Carper D.;
RL Submitted (MAR-1995) to the ENBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;
RA Daoudal S., Berger M., Pailhoux E., Tournaire C., Veyssiere G.,
Jean C.;
RL Submitted (JUN-1995) to the ENBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Ola;
RX MEDLINE=98153248; PubMed=9485485;
RA McGowan M.H., Iwata T., Carper D.A.;
RT "Characterization of the mouse aldose reductase gene and promoter in a lens epithelial cell line."
RN Mol. Vision 4:2-2 (1998).
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RX Ho H.T.B., Jenkins N.A., Copeland N.G., Gilbert D.J., Winkles J.A.,
Louie A.W.Y., Lee F.K., Chung S.S.M., Chung S.K.;
RT "Comparisons of genomic structures and chromosomal locations of the mouse aldose reductase and aldose reductase-like genes."
RN Eur. J. Biochem. 259:726-730 (1999).
[6]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=99160426; PubMed=10049784;
RA Li H., Nobukuni Y., Gui T., Yabe-Nishimura C.;
RT "Characterization of genomic regions directing the cell-specific expression of the mouse aldose reductase gene."
RN Biochem. Biophys. Res. Commun. 255:759-764 (1999).
[7]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smal M.A.,
Schnerch A., Schein J.E., Jones S.J.M., Matra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[8]
RP FUNCTION: Catalyzes the NADPH-dependent reduction of a wide variety of carbonyl-containing compounds to their corresponding alcohols with a broad range of catalytic efficiencies.
[9]
RP CATALYTIC ACTIVITY: Aldol + NAD(P)(+) = aldose + NAD(P)H.
[10]
RP SUBUNIT: Monomer.
[11]
RP SUBCELLULAR LOCATION: Cytoplasmic.
[12]
RP TISSUE SPECIFICITY: Abundant in the testis, skeletal muscle and kidney.
[13]
RP SIMILARITY: Belongs to the aldo/keto reductase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

```



```

GN P100/11E.
OS Leishmania major.
OC Eukaryota; Eukaryozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=89139503; PubMed=2918000;
RA Samaras N., Spichill T.W.;
RT "The developmentally regulated p100/11E gene of Leishmania major shows homology to a superfamily of reductase genes.";
RL J. Biol. Chem. 264:4251-4254(1999).
CC -!- DEVELOPMENTAL STAGE: P100/11E abundance is markedly elevated in promastigotes relative to amastigotes.
CC -!- SIMILARITY: Belongs to the aldo/keto reductase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J04483; AAA7350.1; .
CC PIR; A32950; A32950.
CC HSSP; P06632; LHW6.
CC InterPro; IPR001395; Aldo/ket_red.
CC PRINTS; PF00248; aldo_ket_red; 1.
CC PROSITE; PS00069; ALDOKETREDUCTASE.
CC PRODOM; PD000288; Aldo/ket_red; 1.
CC PROSITE; PS00062; ALDOKETO_REDUCTASE_2; 1.
CC PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.
CC PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
CC Oxidoreductase.
FT ACT SITE 54
FT ACT SITE 54 HYDROGEN-BOND DONOR (BY SIMILARITY).
SQ SEQUENCE 284 AA; 54 82PFEB17F7EDC6CC CRC64;
Query Match 28.38; Score 489.5; DB 1; Length 284;
Best Local Similarity 36.68; Pred. No. 2.9e-31;
Matches 115; Conservative 54; Mismatches 100; Indels 45; Gaps 10;
Qy 8 TSLNGKIPGVGFTFASGSKGE-TYVAVTTALKTGTHLDCAWYLYNEGEVGEIGDF 66
Db 10 TSLNGKVPQFGLGWQSPA--GEVTENAVNWCAGYRHIDTAALYKNEESVGAGLR-- 65
Qy 67 LKENPSVKREDIFVCTKWVHLHYEDVLSIDSLKRLGLDYVDFLHWPTAAEKNGQ 126
Db 66 ---ASGVREDVFITTKLWNTTEQGYSTLAAFEESKQLGVYDLYLIHWP-----R 115
Qy 127 GEPIGPGKYLKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTADLEKMSKFAKVM 186
Db 116 GKDLSEKGYL-----DSWRAFEQLYKEKKVRAIGVSNFHLHLEDVLANCTVT 166
Qy 187 PHAQIIEHPFLPNEELVQVCFKSNIMPVAYSPIGSONQVPTTGERSVENKTNLEIAEK 246
Db 167 PMNQVELHPLNQAOLRAFCDAKQIKVEAWSPLGQ-----GKLIS-NPILSAIGAKY 218
Qy 247 GNTLAQVLIANGVRGVVLPKSNPKRIESEN--FKSIELSDADFAINAVAKGRHFRFV 304
Db 219 NKTRAAQVILRNWIKNLIKTIKPSVHRERIEENADIFDFELGAEDVMSIDA-----L 269
Qy 305 NMKDTFGYDVWPEE 318
Db 270 NTNSRYGPD--PDE 281

```

RESULT 10

ALDR_PIG

ID ALDR_PIG

AC P80276;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

DE Aldo reductase (EC 1.1.1.21) (AR) (Aldehyde reductase).
GN AKR1B1 OR ALR2.
OS Sus scrofa (Pig).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN (1)
RP SEQUENCE, MASS SPECTROMETRY, AND DISULFIDE BOND.
RX TISSUE=Lens;
RX MEDLINE=94109388; PubMed=8281941;
RA Jaquinod M., Potier N., Klarskov K., Reyman J.-M., Sorokine O.,
RA Kieffer S., Barth P., Andrianomanga V., Biellmann J.-F.,
RA van Dorselaer A.;
RT "Spectrometry of pig lens aldoase reductase and electrospray mass spectrometry of non-covalent and covalent complexes.";
RL Eur. J. Biochem. 218:893-903(1993).
RN (2)
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RX MEDLINE=93263021; PubMed=8493902;
RA Kubiseki T.J., Green N.C., Flynn T.G.;
RT "Location of an essential arginine residue in the primary structure of pig aldoase reductase.";
RL Adv. Exp. Med. Biol. 328:259-265(1993).
RN (3)
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=92131138; PubMed=1734286;
RA Rondeau J.-M., Tete-Favrier F., Podjarny A., Reyman J.-M., Barth P.,
RA Biellmann J.-F., Moras D.;
RT "Novel NADPH-binding domain revealed by the crystal structure of aldoase reductase.";
RL Nature 355:469-472(1992).
RN (4)
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=97341224; PubMed=9195881;
RA Urzhumtseva L., Tete-Favrier F., Podjarny A.D., Moras D.;
RA "A 'specificity' pocket inferred from the crystal structures of the complexes of aldoase reductase with the pharmacologically important inhibitors tolrestat and sorbinil.";
RT Structure 5:601-612(1997).
RL -!- FUNCTION: Catalyzes the NADPH-dependent reduction of a wide variety of carbonyl-containing compounds to their corresponding alcohols with a broad range of catalytic efficiencies.
CC -!- CATALYTIC ACTIVITY: Alditol + NAD(P)(+) = aldose + NAD(P)H.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- MASS SPECTROMETRY: MW=35778; MW_ERR=3; METHOD=Electrospray.
CC -!- SIMILARITY: Belongs to the aldo/keto reductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L14950; AAA30989.1; .
CC EMBL; U46065; AAC48515.1; .
CC PIR; A59021; A59021.
CC PDB; 1DLA; 30-APR-94.
CC PDB; 1AH0; 15-APR-98.
CC PDB; 1AH3; 15-APR-98.
CC PDB; 1AH4; 15-APR-98.
CC PDB; 1EKO; 10-MAY-00.
CC InterPro; IPR001395; Aldo/ket_red.
CC Pfam; PF00248; aldo_ket_red; 1.
CC PRINTS; PR00069; ALDOKETREDUCTASE.
CC PRODOM; PD000288; Aldo/ket_red; 1.
CC PROSITE; PS00062; ALDOKETO_REDUCTASE_2; 1.
CC PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.
CC PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.

```

```

DR PRINTS; PRO0069; ALDOXETREDTASE.
DR PRODOM; PD000288; Aldo/ket red; 1.
DR PROSITE; PS00798; ALDOXETO REDUCTASE_1; 1.
DR PROSITE; PS00062; ALDOXETO REDUCTASE_2; 1.
DR PROSITE; PS00063; ALDOXETO REDUCTASE_3; 1.
KW Oxidoreductase; NADP.
FT ACT SITE 48 HYDROGEN-BOND DONOR (BY SIMILARITY).
SQ SEQUENCE 310 AA; 34905 MW; DECCAD94BF3E6D9F CRC64;

Query Match 29.0%; Score 501.5; DB 1; Length 310;
Best Local Similarity 34.8%; Pred. No. 3.7e-32;
Matches 110; Conservative 73; Mismatches 118; Indels 15; Gaps 8;

Qy 6 TFTLSNGVKIPGVCFGTFASEGSGETYATVTTALKTGYHLDCAWYVNLNCEGVGEGIRD 65
Db 3 TVTLSSGYEPMVIGLWRL--KDELKEVILNAIKIGYRHFDCAAHYKSEADVGEALAE 60
Qy 66 FLKENPVQGEIDFVCTKQVNHLYEDVLSDDSLKRLGLDYVDMFLVHWPAAEKNG 125
Db 61 AFKGTG-LVKEELPITTKWNSDGH--VVEACKNSLEKLIQIDYLDLYLVHYEMPTKHA 117
Qy 126 QGEPK--IGPDGKVIILKDTPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKSKFA 193
Db 118 IGTASLGED--KVLDDIVTISLQQTWEGVEKTVSLGLVRSGLSNYELFLTRDCLAYS 175
Qy 184 KMPHANQIEIHPFLPNEELVOYCFSKNIMPVAYSPLGSQNVPTTGERSV--ENKTLNE 241
Db 176 KIKFAVSQFTHPYFQDRLVKFCMKHGVLPHTATPLGGAANKDMFGSVSPLDLDPVND 235
Qy 242 IAEKGNLTALQVILANGRLRGVYVLPKSNPKRIESNFKSIE--LSDADFEAINAVAKGR 299
Db 236 VAKYGYKSAQICLRWGIQRTAVIPKSSKIQRLENLEVLQFQSDQMQLIYSI--DR 293
Qy 300 HRFVNMKDTFGYDVW 315
Db 294 KYRISLPSKWTGLDVI 309

RESULT 8
ALDR RAT STANDARD; PRT; 315 AA.
AC P07943;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aldose reductase (EC 1.1.1.21) (AR) (Aldehyde reductase).
GN AKR1B1 OR ALDR1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Lens;
RA MEDLINE=87276556; PubMed=3111886;
RA Carper D., Nishimura C., Shiohara T., Dietzchold B., Wistow G.,
RA Craft C., Kador P., Kinoshita J.H.;
RT "Aldose reductase and p-cristallin belong to the same protein
RT superfamily as aldehyde reductase."
RL FEBS Lett. 220:209-213(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92084118; PubMed=1748296;
RA Graham C.E., Szpirer C., Levan G., Carper D.;
RT "Characterization of the aldose reductase-encoding gene family in
RT rat."
RL Gene 107:259-267(1991).
RN [3]
RP SEQUENCE OF 155-168 AND 204-209.
RC TISSUE=Astrocytes;
RX MEDLINE=96007849; PubMed=7498172;
RA Laeng P., Bouillon P., Raupenot L., Labourdette G.;
RT "Long-term induction of an aldose reductase protein by basic
RT

```

```

RT fibroblast growth factor in rat astrocytes in vitro."
RL Electrophoresis 16:1240-1250(1995).
CC FUNCTION: Catalyzes the NADPH-dependent reduction of a wide
CC variety of carbonyl-containing compounds to their corresponding
CC alcohols with a broad range of catalytic efficiencies.
CC CATALYTIC ACTIVITY: Alditol + NAD(P)(+) = aldose + NAD(P)H.
CC SUBUNIT: Monomer.
CC SUBCELLULAR LOCATION: Cytoplasmic.
CC SIMILARITY: Belongs to the aldo/keto reductase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; X05884; CAA29308.1; -.
CC EMBL; M60322; AAA40721.1; -.
CC PIR; A60603; A60603.
CC HSP; P15121; 2AQO.
CC InterPro; IPR001395; Aldo/ket_red.
CC Pfam; PF00248; aldo_ket_red; 1.
CC PRINTS; PRO0069; ALDOXETREDTASE.
CC PRODOM; PD000288; Aldo/ket red; 1.
CC PROSITE; PS00062; ALDOXETO REDUCTASE_2; 1.
CC PROSITE; PS00063; ALDOXETO REDUCTASE_3; 1.
CC PROSITE; PS00798; ALDOXETO REDUCTASE_1; 1.
KW Oxidoreductase; NADP; Acetylation.
FT INIT MET 0 BY SIMILARITY.
FT NP BIND 9 18 NADP (POTENTIAL).
FT ACT SITE 48 48 HYDROGEN-BOND DONOR (BY SIMILARITY).
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 315 AA; 35666 MW; 572941A154BC1202 CRC64;

Query Match 28.9%; Score 499.5; DB 1; Length 315;
Best Local Similarity 39.4%; Pred. No. 5.4e-32;
Matches 121; Conservative 51; Mismatches 106; Indels 29; Gaps 10;

Qy 9 LSNGVKIPGVGFTFASGSGKETVAVTTALKTGYRHLDCAWYVNLNCEGVGEGIRDFLK 68
Db 6 LNNGYKPTLGLTWKS--PQGVTEAVKVAIDGYRHIDCAQVYQNEKEVGVALQEKLK 63
Qy 69 ENPSVKREDIFVCTKQVNHLYEDVLSDDSLKRLGLDYVDMFLVHWPAAEKNGOGE 128
Db 64 EQ-VVKKQDLFTVSKLWCTFFHQSMVKGACQKTLSDQLQDLYLDLIHWPTGF----- 115
Qy 129 PKIGPDGKVIILKDTENPEP-----TWRAVEKIYEDRKARSIGVSNWTIADLEKM-- 179
Db 116 -KPGPD--YFPL-DASGNVPSDDTDFVTWTAMEQLVDEGLVKAIGVSNFNPQLIERILN 171
Qy 180 SKFAKVMPHANQIEIHPFLPNEELVOYCFSKNIMPVAYSPLGSQNVPTTGERSV--ENK 237
Db 172 KEGLYKPAVNGIECHPVLTOEKLIEYCHGIVVATYSPILGSPPDRPAKPEDPSLLEDP 231
Qy 238 TLNEIAEKGNTLAQVILANGRLRGVYVLPKSNPKRIESNEK--SIELSDADFEAINAV 295
Db 232 RIKEAAKYNKTTAGVILRFPIQRNLVWIPKSVTPARIAENPKVDFELSDMATL--L 289
Qy 296 AKGRHFR 302
Db 290 SYNRRWR 296

RESULT 9
P100 LEIMA STANDARD; PRT; 284 AA.
ID P100 LEIMA
AC P22045;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable reductase (EC 1.1.-.-).

```

```

RESULT 6
ID _ALD1_MOUSE STANDARD; PRT; 315 AA.
AC P21300;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Aldose reductase-related protein 1 (EC 1.1.1.21) (AR) (Aldehyde
DE reductase) (VAS deferens androgen-dependent protein) (MVDP)
DE (Aldo-keto reductase family 1 member B7).
DE AKR1B7 OR AVDP.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
[1]
RN _SEQUENCE FROM N.A.
RC TISSUE=Vas deferens;
RX MEDLINE=92345100; PubMed=1637719;
RA Pailhoux E.A., Veyssi re G.M., Fabre S., Tournaire C., Jean C.G.;
RT "The genomic organization and DNA sequence of the mouse vas deferens
RT androgen regulated protein gene.";
RN J. Steroid Biochem. Mol. Biol. 42:561-568(1992).
[2]
RN _SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Vas deferens;
RX MEDLINE=91060614; PubMed=2123194;
RA Pailhoux E.A., Martinez A., Veyssi re G.M., Jean C.G.;
RT "Androgen-dependent protein from mouse vas deferens. cDNA cloning and
RT protein homology with the aldo-keto reductase superfamily.";
RN J. Biol. Chem. 265:19932-19936(1990).
CC -!- FUNCTION: The role of MVDP in sperm maturation and storage may be
CC related to its potential capacity to produce fructose, or MVDP may
CC play an osmoregulatory role by producing sorbitol.
CC -!- CATALYTIC ACTIVITY: Alditol + NAD(P)(+) = aldose + NAD(P)H.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- INDUCTION: Castration resulted in a marked decrease in the level
CC of the mRNA coding for the protein, whereas administration of
CC testosterone to castrated males resulted in a marked increase.
CC -!- SIMILARITY: Belongs to the aldo/keto reductase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M81448; AAA39774.1; -.
DR EMBL; J05663; AAA39773.1; -.
DR PIR; A37990; A37990.
DR HSP; P45377; 1FRB.
DR MGD; MG1:101918; AKR1B7.
DR InterPro; IPR001395; Aldo/ket_red.
DR PRINTS; PF00248; aldo_ket_red; 1.
DR PROSITE; PS00069; ALDOKETREDTASE.
DR PROSITE; PS000288; Aldo/ket_red; 1.
DR PROSITE; PS00062; ALDOKETO_REDUCTASE_2; 1.
DR PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.
DR PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
DR PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
DR OXIDOREDUCTASE; NADP.
DR INIT_MET 0 0 BY SIMILARITY.
DR ACT_SITE 48 48 HYDROGEN-BOND DONOR (BY SIMILARITY).
DR FT CONFLICT 93 93 Q->D (IN REF. 2).
DR _SEQUENCE 315 AA; 35657 MW; EA3D528294AF8542 CRC64;
Query Match 29.1%; Score 503; DB 1; Length 315;
Best Local Similarity 38.7%; Pred.No. 2.8e-32;
Matches 123; Conservative 58; Mismatches 121; Indels 16; Gaps 9;
6 TFT-LSNGVKIPGVGFGTFSESGKSGTGTAVTALTGTGRHLDCAWYLNCEGEVGEIR 64

```

QY

```

DR GO; GO:0016491; F:oxidoreductase activity; IDA.
DR GO; GO:0019566; P:arabinose metabolism; IMP.
DR InterPro: IPR001395; Aldo/ket red.
DR Pfam: PF00248; aldo ket red; 1.
DR PRINTS: PR00069; ALDKETREDTASE.
DR ProDom: PD000288; Aldo/ket_red; 1.
DR PROSITE; PS00062; ALDOKETO_REDUCTASE_2; 1.
DR PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.
DR PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
KW Oxidoreductase.
FT ACT SITE 56 56 HYDROGEN-BOND DONOR (BY SIMILARITY).
SQ SEQUENCE 312 AA; 34755 MW; 078F92FCACECC70E CRC64;

Query Match 32.1%; Score 555; DB 1; Length 312;
Best Local Similarity 40.1%; Pred. No. 2.3e-36;
Matches 128; Conservative 60; Mismatches 97; Indels 34; Gaps 9;

QY 2 SNGTFTLSNGVKTGPGVGFGTFASESGSGKGETYATVTTALKTGYRHLDCAWYVINEGEVGE 61
DB 7 NSSATLKNTGASIPVLGFGTWRSVDNNG--YHSVIALKAGYRHDAAYIYLNBEVGR 64
QY 62 GIRDFLKENPSVKREDIFVCTKWNHLHRYEDVLMSIDDSLKRLGLDYVDMFLVHPPIAA 121
DB 65 AIKD-----SGVPREEIFITKLWGTEQR--DPEAALNLSKRLGLDYVDMFLVHPVPL 117
QY 122 EKNQCGEPKTPGDCGYVLKDLTE-----NPEPTWRAMEKIYEDRKASIGVSNW 171
DB 118 KTDV-----TDGNVICPTLEDGTVDITDKWNFIKTWELMQELPKTKTKAVGVSNF 171
QY 172 TIADLRK--SKFAKVMPHANQIIEHPFLNEELVQVCFSKNTMPVAYSPLGSGNQVPTT 229
DB 172 SINNIKELLESPPNKKVPATNIEIHPLLPQDELIATCEKGIIVEAYSFFGSAN-APLL 230
QY 230 GERVSEKNTLNEIAEKGNVLAQVLIANGLRGVVPLPKSSNPKRIESNPKSIELSDADF 289
DB 231 KE-----QAIDWAKKEGVEPAQLIISWSIQRGVVLAKSVNPERIVSNFKITLPEDDF 285
QY 290 EAINAVAKGRHF-RFVNKK 307
DB 286 KTIISNLKSVHGKKEVDYK 304

RESULT 5
ARAI_YEAST STANDARD; PRT; 344 AA.
AC P39115;
DT 01-OCT-1994 (Rel. 30, Created)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE D-arabinose dehydrogenase [NAD(P)+] heavy chain (EC 1.1.1.117).
GN ARA1 OR YBR149W OR YBR1127.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OK NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RA Entian K.-D., Koetter P., Rose M., Becker J., Grey M., Li Z.,
RA Niesenmann E., Schenk-Groeninger R., Servos J., Wehner E.,
RA Wolter R., Brendel M., Bauer J., Braun H., Derr K., Duesterhus S.,
RA Gruenbein R., Hedges D., Kieseppa P., Korol S., Krebs B., Proft M.,
RA Siegers K., Baur A., Boles E., Miosga T.,
RA Schaaff-Gerstenschlaeger I., Zimmermann F.K.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91171289; PubMed=2005616;
RA Martinez-Soriano J.P., Wong W.M., van Ryk D.I., Nazar R.N.;
RT "A widely distributed 'CAT' family of repetitive DNA sequences.";
RL J. Mol. Biol. 217:629-635(1991).
RN [3]
RP CHARACTERIZATION, AND SEQUENCE OF 7-20.

```

```

RX MEDLINE=99117109; PubMed=9920381;
RA Kim S.T., Huh W.K., Lee B.H., Kang S.O.;
RT "D-arabinose dehydrogenase and its gene from Saccharomyces
RT cerevisiae";
RL Biochim. Biophys. Acta 1429:29-39(1998).
CC -!- FUNCTION: Catalyzes the oxidation of D-arabinose, L-
CC fucose and L-galactose in the presence of NADP+.
CC -!- CATALYTIC ACTIVITY: D-arabinose + NAD(P) (+) = D-arabinono-1,4-
CC lactone + NAD(P)H.
CC -!- SUBUNIT: Heterodimer of a heavy chain and a light chain.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- MISCELLANEOUS: Exhibits maximum activity at pH 10.0 and around 30
CC degrees Celsius.
CC -!- SIMILARITY: Belongs to the aldo/keto reductase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; Z36018; CAAB5107.1; -.
DR EMBL; W95580; AAB35037.1; -.
DR EIR; S46020; S46020.
DR HSSP; P06632; 1HW6.
DR GerMOnline; 138692; -.
DR SGD; S000353; ARA1.
DR GO; GO:0045290; F:D-arabinose 1-dehydrogenase [NAD(P)] activity; IDA.
DR GO; GO:0005975; P:carbohydrate metabolism; IDA.
DR InterPro: IPR001395; Aldo/ket_red.
DR Pfam; PF00248; aldo_ket_red; 1.
DR PRINTS; PR00069; ALDKETREDTASE.
DR ProDom; PD000288; Aldo/ket_red; 1.
DR PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
DR PROSITE; PS00062; ALDOKETO_REDUCTASE_2; 1.
DR PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.
KW Oxidoreductase; NAD.
FT ACT SITE 71 71 HYDROGEN-BOND DONOR (BY SIMILARITY).
FT CONFLICT 56 62 AAIKAGY -> LQSKLDN (IN REF. 2).
FT CONFLICT 71 73 YET -> SR (IN REF. 2).
SQ SEQUENCE 344 AA; 38883 MW; F9B6D9333B18FECC CRC64;

Query Match 31.6%; Score 547.5; DB 1; Length 344;
Best Local Similarity 37.4%; Pred. No. 1e-35;
Matches 122; Conservative 58; Mismatches 91; Indels 55; Gaps 8;

QY 7 FTLSNGVPIPGVGFGTFASESGSGKGETYATVTTALKTGYRHLDCAWYVINEGEVSGIRD 66
DB 25 FSLNGVRIPALGLGTANPHEKLAETKQAKAAIKAGYRHDTAWAYETEPFVGAIKEL 84
QY 67 LKENSVMKREDIFVCTKWNHLHRYEDVLV-----SIDDSLKRLGLDYVDMFLVHPPIAA 121
DB 85 L-EDGSIKREDLFTTKW-----PVLWDEVRSLNESLKALGLEVYDVLQLQHWPLCF 136
QY 122 EK-----NQQGEPIKPGDQKGVILKDLTENPEPTWRAMEKIY---EDRKA 163
DB 137 EKIKDPKGISGLVKTPVDDSGKTYAAGDYL-----ETVKQLEKIYLDPNDRHV 186
QY 164 RIGVSNVTIADLEKMSKFAKMPHANQIIEHPFLNEELVQVCFSKNTMPVAYSPLGSG 223
DB 187 RAIGVSNFSIEYLERLIEKRCVKPTVNVQVETHPLPQMELEKRCFCFMDILLTAYSPGSH 246
QY 224 N----QVPTTGERVSEKNTLNEIAEKGNVLAQVLIANGLRGVVPLPKSSNPKRIESNF 279
DB 247 GAPNLKIP-----LVKKLAKYNTGNDLLISVHIHQGTIVIRSLNPNVRISSI 296
QY 280 KSIELSDADFAINAVAKGRHFRFVN 305
DB 297 EFASLTQDELQELNDGKYPVRFID 322

```



```

DR EMBL; X90518; CAA62107.1; -.
DR EMBL; X94335; CAA64040.1; -.
DR EMBL; X96740; CAA65512.1; -.
DR EMBL; Z75028; CAA99318.1; -.
DR PIR; S22846; S22846.
DR HSP; P06632; 1HW6.
DR Germonline; 143708; -.
DR SGP; S0005645; GCY1.
DR GO; GO:0004033; P:aldo-keto reductase activity; IDA.
DR GO; GO:0009551; P:salinity response; IEA.
DR InterPro; IP001395; Aldo/ket red.
DR Pfam; PF00248; aldo_ket_red; I.
DR PRINTS; PR00069; ALDOKETREDASE.
DR ProDom; PD00288; Aldo/ket_red; 1.
DR PROSITE; PS00062; ALDOKETO_REDUCTASE_2; 1.
DR PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.
DR PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
KW Oxidoreductase.
FT ACT SITE 56
SQ SEQUENCE 312 AA; 35079 MW; 77B9D31C228334N4 CRC64;

Query Match 33.2%; Score 574.5; DB 1; Length 312;
Best Local Similarity 40.8%; Pred. No. 6.8e-38;
Matches 127; Conservative 54; Mismatches 85; Indels 45; Gaps 7;

QY 5 KTFSLNGVKIPGVGCTFASEGSKGETYTAVTALKTGYRHLDCAWYLYNEGEVGEIR 64
Db 10 KLSLNTGALPQGLGTGWSK--ENDAYKAVLTALKDGRHDTAAIYRNDQVQQA 67
QY 65 DFLKENPSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGDDYVDMFLVHWP 124
Db 68 D-----SGVPRFEIVTTKLWCTQHPEV--ALDQSLKELGDDYVDMFLVHWP 117
QY 125 GQGEPKIPGDKYVILKDLTENP-----PTWRAMEKIYEDRKASIGV 168
Db 118 -----DPAYIKNEDILSVPTKDGSRADITNNFKITWELMOELPKTKTAVGV 168
QY 169 SNWTIADLRK--SKFAKVMPHANQIHPFLPNEELVQYCFKSNIMPVAYSPLGSONQV 226
Db 169 SNFSINNKLKLLASQGNKLTAAQVVEIHPQLQDELINCKSKGIWEAYSPLGS---- 224
QY 227 PTTGERVSEKNTLNEIAEKGNLTAAQVLIAGLRGVVILPKSNPKRIBSNFKSLSD 286
Db 225 --TDAPLKEFVILEIAKKNVQPGHVVISWHVQRGVYVLPKSNPDINKRKFTLST 282
QY 287 ADPEAINAVAK 297
Db 283 EDPEAINNISK 293

RESULT 2
YDG7 SCHPO STANDARD; PRT; 321 AA.
AC Q10494;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable oxidoreductase C26F1.07 in chromosome I (EC 1.1.1.1).
GN SPAC26F1.07.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgours J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

```

```

RA Holroyd S., Hornaby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Coffeau A., Cadieu E., Dreano S., Clout S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -!- SIMILARITY: Belongs to the aldo/keto reductase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z73100; CAA97364.1; -.
DR PIR; T38413; T38413.
DR HSP; P14550; 2ALR.
DR GensDB_Spombe; SPAC26F1.07; -.
DR InterPro; IPR001395; Aldo/ket_red.
DR Pfam; PF00248; aldo_ket_red; 1.
DR PRINTS; PR00069; ALDOKETREDASE.
DR ProDom; PD00288; Aldo/ket_red; 1.
DR PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
DR PROSITE; PS00062; ALDOKETO_REDUCTASE_2; 1.
DR PROSITE; PS00063; ALDOKETO_REDUCTASE_3; FALSE_NEG.
KW Hypothetical protein; Oxidoreductase.
FT ACT SITE 60
SQ SEQUENCE 321 AA; 36192 MW; 6B03CB14A726FC0B CRC64;

Query Match 32.8%; Score 568; DB 1; Length 321;
Best Local Similarity 43.0%; Pred. No. 2.3e-37;
Matches 131; Conservative 59; Mismatches 89; Indels 26; Gaps 10;

QY 7 FTLSNGVKIPGVGCTFASEGSKGETYTAVTALKTGYRHLDCAWYLYNEGEVGEIRDF 66
Db 16 FTLDGSKIPGLGLGTWSE--PNQTKAVKTAQYGYRHIDAAIYGNEDVGDGIKE- 72
QY 67 LKENPSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGDDYVDMFLVHWP 126
Db 73 ----SGVPRKDIWVTSKLVKNAHAPEAVFKALETKDLKLDYLDYLIHVPVSF-KTGE 127
QY 127 GEPKICPDGKYVILKDLTENP-EPTWRAMEKIYEDRKARSIGVSNWTIADLRKSKFAKY 185
Db 128 DKFPKDKGNLIYK-----NPIETWKAVEKLETKGVKHGLHSNFDNLERILKVAKV 183
QY 186 MPHANQIHPFLPNEELVQYCFKSNIMPVAYSPLGSON-----QVPTGSRVSENKTLN 240
Db 184 KPAVHQMLHPFLPQTEFEVEXKHKLGIHVTAYSPFGNQNTIYESKIP----KLIEHETIQ 239
QY 241 ELAE--KG--GNTLAQVLIAGLRGVVILPKSNPKRIBSNFKSLSDADPEAINAVAKG 298
Db 240 KIAKSGEGVGTATVATVATRGTSVLPKSNVNEQRIKSNFKYPLTKDWDENSI--G 297
QY 299 RHFRF 303
Db 298 IRARF 302

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 13:59:56 ; Search time 18 Seconds
(without alignments)
940.156 Million cell updates/sec

Title: US-10-004-115B-1

Perfect score: 1731

Sequence: 1 MSNGKFTLSNGVKIGVGF.....MKDTFGYDWPBETAKULSA 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	574.5	33.2	312	1 GCY YEAST	P14065 saccharomyc
2	568	32.8	321	1 YDG7 SCHPO	Q10494 schizosacch
3	559.5	32.3	322	1 ALDM SPOSA	P27600 sporobolomyc
4	555	32.1	312	1 YPR1 YEAST	Q12458 saccharomyc
5	547.5	31.6	344	1 ARA1 YEAST	P38115 saccharomyc
6	503	29.1	315	1 ALD1 MOUSE	P21300 mus muscullu
7	501.5	29.0	310	1 S6PD MALDO	P28475 malus domes
8	499.5	28.9	315	1 ALDR RAT	P07943 rattus norv
9	489.5	28.3	284	1 P100 LEIMA	P22045 leishmania
10	488.5	28.2	315	1 ALDR PIG	P80276 sus scrofa
11	488.5	28.2	315	1 ALDR RABIT	P15122 oryctolagus
12	487.5	28.2	315	1 ALDR MOUSE	P45376 mus muscullu
13	487	28.1	316	1 AKBA HUMAN	O60218 homo sapien
14	486.5	28.1	315	1 ALDR HUMAN	P15121 homo sapien
15	486	28.1	318	1 XYLI PACTA	P78736 pachysolen
16	486	28.1	323	1 DHBS MOUSE	P70494 mus muscullu
17	485	28.0	315	1 ALD2 MOUSE	P45377 mus muscullu
18	480.5	27.8	301	1 AKR1 MOUSE	Q9dct1 mus muscullu
19	478	27.6	315	1 ALDR BOVIN	P16116 bos taurus
20	478	27.6	324	1 AKAI MOUSE	Q9j116 mus muscullu
21	477	27.6	324	1 AKAI HUMAN	P14550 homo sapien
22	475.5	27.5	323	1 PE2R RABIT	P80508 oryctolagus
23	473	27.3	324	1 AKAI RAT	P51635 rattus norv
24	467	27.0	324	1 AKAI PIG	P50378 sus scrofa
25	461	26.6	326	1 AKD1 HUMAN	P51857 homo sapien
26	457	26.4	323	1 AKC2 HUMAN	P52895 h aldo-keto
27	456	26.3	295	1 MORA PSBPU	Q02198 pseudomonas
28	454	26.2	323	1 AKC1 HUMAN	Q04828 h aldo-keto
29	450.5	26.0	323	1 PGFS BOVIN	P05960 bos taurus
30	447.5	25.9	323	1 AKC3 HUMAN	P42330 h aldo-keto
31	445.5	25.7	323	1 PGF2 BOVIN	P52897 bos taurus
32	445	25.7	327	1 GRE3 YEAST	P38715 saccharomyc
33	443.5	25.6	323	1 DBX3 BOVIN	P52898 bos taurus

34 439 25.4 329 1 XYLI_KULJA
35 438.5 25.3 323 1 AKC4_HUMAN
36 437 25.2 326 1 AKD1_RAT
37 435.5 25.2 323 1 CRO_RANCA
38 432.5 25.0 323 1 CRO_RANTE
39 432 25.0 323 1 ALDR_HORVU
40 430.5 24.9 323 1 PE2R_RAT
41 420.5 24.3 318 1 XYLI_PIGST
42 419.5 24.2 322 1 DIDH_RAT
43 414.5 23.9 321 1 DTDH_MUGMU
44 414 23.9 275 1 DKGA_ECOLI
45 413 23.9 282 1 YJ66_YEAST

ALIGNMENTS

RESULT 1
GCY YEAST
ID GCY YEAST STANDARD; PRT; 312 AA.
AC P14065;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE GCY protein [EC 1.1.1.-]
GN GCY1 OR GCV OR YOR120W OR O31567 OR YOR3269W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D273-10B;
RX MEDLINE=89005653; PubMed=2901985;
RA Oechsner U., Magdolen V., Bandlow W.;
RT "A nuclear yeast gene (GCY) encodes a polypeptide with high homology to a vertebrate eye lens protein.";
RL FEBS Lett. 238:123-128(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=97060020; PubMed=8504341;
RA Wiemann S., Rechmann S., Benes V., Voss H., Schwager C., Vlcek C., Stegemann J., Zimmermann J., Effle H., Paces V., Ansoerge W.;
RT "Sequencing and analysis of 51 kb on the right arm of chromosome XV from Saccharomyces cerevisiae reveals 30 open reading frames.";
RL Yeast 12:281-288(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97344368; PubMed=9200815;
RA Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C., Schwager C., Paces V., Sander C., Ansoerge W.;
RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV.";
RL Yeast 13:655-672(1997).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 25657;
RA Bandlow W.;
RT Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Unknown function which seems to be not essential.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE FAMILY.
CC VERY SIMILAR TO YEAST YPR1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
CC
CC EMBL; X13228; CAA31615.1; -.

Search completed: May 28, 2004, 14:06:18
Job time : 21 secs

```
Db      3 TYQLNNGVEIPVLGFGTFKAKDGE-EAYRAVLEALKAGYRHIDTAAYIQNEESVGOAID 61
QY      66 FLKENSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYVDMPFLVHPPIAAEKNG 125
Db      62 -----SGVPREEMFVTTKLWNSQQTVEQTRQALEKSIKELGLDYLGLYLIHWP----- 109
QY      126 QGEPKIGPDGKYVILKDLTENPEP-----TWAMEKIYEDRKARSIGVSNWT 172
Db      110 -----NPKPLRENDANKTRNAEVVRAMEDLYQEGKIRAIQVSNFL 149
QY      173 IADLEKMSKFAKVMPHANQIEHPFLPNBELVOYCFKSNIMPVAYSPGLSQNVPTTGER 232
Db      150 PHLDALLEATATVPAVNVRLAPGVYQDVAYCREKGILLLEAWGPFQO-----GE- 201
QY      233 VSENKTLNEIAEKGNTLAQVLAWLRGCVVLPKSSNPKRIESNFK--SIELSDADFE 290
Db      202 LFDSKQVQEIQAANHGKVAQIALAWSLAEGFLPLPKSVTTSRIQANLDCFGIELSHEERE 261
QY      291 AINAVA 296
Db      262 TLKTIA 267

RESULT 15
C98038
conserved hypothetical protein spr1332 [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C;Accession: C98038
R;Hoskins J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
y, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: C98038
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-280 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAL00136.1; PID:g15458977; GSPDB:GN00174
C;Genetics:
A;Gene: spr1332
C;Superfamily: aldehyde reductase

Query Match      29.8%; Score 516; DB 2; Length 280;
Best Local Similarity 35.9%; Pred. No. 7.8e-33;
Matches 110; Conservative 57; Mismatches 83; Indels 56; Gaps 7;

QY      6 TFTLSNGVKIPGVGTFSESGSGKGETYATVTTALKTGYRHLDCAWYVINEGEGVGEIRD 65
Db      3 TYQLNNGVEIPVLGFGTFKAKDGE-EAYRAVLEALKAGYRHIDTAAYIQNEESVGOAID 61
QY      66 FLKENSVKREDIFVCTKWNHLHRYEDVLSIDSLKELGLDYVDMPFLVHPPIAAEKNG 125
Db      62 -----SGVPREEMFVTTKLWNSQQTVEQTRQALEKSIKELGLDYLGLYLIHWP----- 109
QY      126 QGEPKIGPDGKYVILKDLTENPEP-----TWAMEKIYEDRKARSIGVSNWT 172
Db      110 -----NPKPLRENDANKTRNAEVVRAMEDLYQEGKIRAIQVSNFL 149
QY      173 IADLEKMSKFAKVMPHANQIEHPFLPNBELVOYCFKSNIMPVAYSPGLSQNVPTTGER 232
Db      150 PHLDALLEATATVPAVNVRLAPGVYQDVAYCREKGILLLEAWGPFQO-----GE- 201
QY      233 VSENKTLNEIAEKGNTLAQVLAWLRGCVVLPKSSNPKRIESNFK--SIELSDADFE 290
Db      202 LFDSKQVQEIQAANHGKVAQIALAWSLAEGFLPLPKSVTTSRIQANLDCFGIELSHEERE 261
QY      291 AINAVA 296
Db      262 TLKTIA 267
```

QY 127 GEPKIGDPGVKILKDTENPBTWRAMEKIYEDRKARSIGVSNWTIADLEKMSKPAKYM 186
Db 119 ---SLMPTPEMLKPDITS---TWKAMEALYSGKARAIYGVSNFSSKLTDLNVARVT 171
QY 187 PHANQIIEHFLPNEELVQYCFKSNIMPVAYSPLGSONQVPTTGE---RVSENKTLNEIA 243
Db 172 PAVNOVECHFPWQOGLHELCKSGVHLSGYSPLGSQSK---GEVRLKVLQNPVITEVA 227
QY 244 EKGNTLAQVLIAGLARGVIVLPKSNPKRIEISNFKSIELSDAD--FEAINAVAKGRHP 301
Db 228 EKLKGTTAQVALRMGLQTHSVLPKSSSGARLKENLDVDFWMSIPEDLFTKFSNIPQASSF 287
QY 302 RFV 304
Db 288 IFL 290

RESULT 12
T09670
absic acid activated protein - alfalfa
C:Species: Medicago sativa (alfalfa)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
R:Deal, M.; Davletova, S.; Oberschall, A.; Meszaros, T.; Miskolczi, P.; Torok, K.; Dudits
submitted to the EMBL Data Library, April 1996
A:Description: Calmodulin inhibitors enhance the expression of abscisic acid-activated g
A:Reference number: T16812
A:Accession: T09670
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-313 <DEA>
A:Cross-references: EMBL:X97606
A:Experimental source: cultivar Regen S; strain RA3
C:Superfamily: aldehyde reductase

Query Match 30.28; Score 523.5; DB 2; Length 313;
Best Local Similarity 37.28; Pred. No. 2.4e-33;
Matches 123; Conservative 61; Mismatches 106; Indels 41; Gaps 10;

QY 5 KFTLSNGVKIPGVGFGTFASEGSKGETYTAVTALKTGYRHLDCAWYLNAGEVGEGR 64
Db 6 KFFQNTGAKIPSVGLGTWQAE--PGVAVKAVITAVGVYRHIDCAEAYKNQSEIGSALK 63
QY 65 DFLKENPSVKREDIFVCTKWNHLHRYEDVLWISDLSKRLGLDYDMFLVHPPIAAEK 124
Db 64 K-LCEDGVKREELWITSKLWCSDHPEDVPKALDKTINDLQLDYLDLYLHWPVSM-KR 121
QY 125 GQGPVKIGDPGVKILKDTENPE---PTWRAMEKIYEDRKARSIGVSNWTIADLEKM 179
Db 122 GTGE-----FMGENDLHDADIPSTWKALGYDSGAKAIGVSNFSTKQLDL 168
QY 180 SKFAKVMPHANQIIEHFLPNEELVQYCFKSNIMPVAYSPLGSONQVPTTGERVSENKTL 239
Db 169 LDVARVPVAVNQVHELHPGWOQAKLHAFCSKGHLSGYSPLGSPGVLKSD---ILKNPVV 225
QY 240 NEIAEKGNTLAQVLIAGLARGVIVLPKSNPKRIEISNFKSIELSDAD----FEAI-- 292
Db 226 KEIAEKLGKTPGQVALRWGLQAGHSVLPKSNRIKKNLDVWISIPEDLFPKFSIKQ 285
QY 293 NAVAKGRHFRVNMKDTFG-----YDVWPEE 318
Db 286 DKLIKGTFF---VNDTVGAPRTIEELWDGE 312

RESULT 13
T48188
aldose reductase-like protein - Arabidopsis thaliana
N:Alternate names: protein F7A7.190
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T48188
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De

eves, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24487
A:Accession: T48188
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-320 <BEV>
A:Cross-references: EMBL:AL161946
A:Experimental source: cultivar Columbia; BAC clone F7A7
C:Genetics:
A:Map position: 5
A:Introns: 48/3; 65/3; 90/2; 119/3; 188/3; 210/3; 291/3
A:Note: F7A7.190
C:Superfamily: aldehyde reductase

Query Match 30.11; Score 521; DB 2; Length 320;
Best Local Similarity 35.61; Pred. No. 3.8e-33;
Matches 117; Conservative 73; Mismatches 105; Indels 34; Gaps 10;

QY 3 NGKFTLSNGVKIPGVGFGTFASEGSKGETYTAVTALKTGYRHLDCAWYLNAGEVGE 62
Db 12 NMESFLLSGHKIPAVGLGTWRS--GSQA-AHAVVTALVEGGYRHLDTAWEGDQREVQCG 69
QY 63 IRDFLENPSVKREDIFVCTKWNHLHRYEDVLWISDLSKRLGLDYDMFLVHPPIAAE 122
Db 70 IKRAM--HAGLERRDLFTYKLMCTELSPERVRPALQNTLKELEYLDLYLHWPFLR 127
QY 123 KNGQGEPIGDPGVKILKDLTE--NPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMSK 181
Db 128 EGASKPPKAG-----DVLDFDMEGWREMNLSKDSLVRNIGVCNFTVKLNKLLG 178
QY 182 FAKVMPHANQIIEHFLPNEELVQYCFKSNIMPVAYSPLGSONQVPTTGERVSENKTLNE 241
Db 179 FAELIPAVCOMEMHFGWENDRILEFCCKNEIH--AYSPLGSQE---GGRDLIHDQTYDR 232
QY 242 IAEKGGNTLAQVLIAGLARGVIVLPKSNPKRIEISNFKSIE--LSDADFEAINAVAKGR 299
Db 233 IAKLNKTPGQILYKWLQRTGTSVIPKSLNPERIKENIKVFDWVPEQDFQALNSITDQK 292
QY 300 HRFVNMKDTF-----GYDVWPEE 318
Db 293 --RVIDGEDLFVNKTEGPFGRSVADLWDHE 319

RESULT 14
C95172
oxidoreductase, aldo/keto reductase family SP1478 [imported] - Streptococcus pneumoniae
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95172
R:Tetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95172
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-280 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75572.1; PID:g14972968; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1478
C:Superfamily: aldehyde reductase

Query Match 29.81; Score 516; DB 2; Length 280;
Best Local Similarity 35.91; Pred. No. 7.8e-33;
Matches 110; Conservative 57; Mismatches 83; Indels 56; Gaps 7;

QY 6 TFTLSNGVKIPGVGFGTFASEGSKGETYTAVTALKTGYRHLDCAWYLNAGEVGEGRD 65

A;Description: Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence.
 A;Reference number: Z14677
 A;Accession: T02543
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-290 <ROU>
 A;Cross-references: EMBL:AC004684; NID:G3236234; PID:G3236259
 A;Experimental source: cultivar Columbia
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A;Reference number: A84420; MUID:20083487; PMID:10617197
 A;Accession: H84796
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-290 <STO>
 A;Cross-references: GB:AE002093; NID:G3236259; PIDN:AAC23647.1; GSPDB:GN00139
 C;Genetics:
 A;Gene: F13M22.27; At2g37770
 A;Map position: 2
 A;Introns: 40/3; 57/3; 83/2; 112/3; 204/3
 C;Superfamily: aldehyde reductase

Query Match 30.6%; Score 529; DB 2; Length 290;
 Best Local Similarity 42.3%; Pred. No. 7.9e-34;
 Matches 124; Conservative 51; Mismatches 98; Indels 20; Gaps 8;
 QY 1 MSNGKT-FTLSNGVKIPGVGFTF-ASEGSGKGTYYTAVTTALKTGYRHLDCAWYLYNEGE 58
 DB 1 MANATTFKLTGAKFVSGLGTQWASPLVGD---AANAAGVIGYRHIDCAIYNEKE 57
 QY 59 VGGIRDFLKENPSVKREDIFVCTKWNHLHRYEDVLMSSIDSLKRLGLDYDMFLVHP 118
 DB 58 IGAVLKK-LPEDRVKREDLFITKLMCTDHDQDPVPEALNRLTKDLQLEVDLYLHP 116
 QY 119 IAAKNGQGPFGDKGYVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEK 178
 DB 117 ARIKGSVG---IKPEML-----LPVDIPSTWAMEALYDSGKARAIGVSNFSTKKLAD 167
 QY 179 MSKFAKYMPHANQIEIHPFLNEELVQYCFKSNIMPVAYSPGSONQVPTTGER--VSEN 236
 DB 168 LLELARVPNAVQVECHPSWRQTKIQEFCKSGVHLSAYSPGSG---TTWLKSDVLKN 224
 QY 237 KTLNEIAEKGNLAQVLIANGRLRGVYVLPKSSNPKRIESNFKSIELSDADF 289
 DB 225 PILMVAEKLKSPQAQVALEWGLQMGHSLVLPKSTNEGRIKENFNVDWSPIDY 277

RESULT 10
 B84599
 hypothetical protein At2g21260 [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C;Accession: B84599
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A;Reference number: A84420; MUID:20083487; PMID:10617197
 A;Accession: B84599
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-309 <STO>
 A;Cross-references: GB:AE002093; NID:G4567261; PIDN:AAD23674.1; GSPDB:GN00139
 C;Genetics:
 A;Gene: At2g21260
 A;Map position: 2
 C;Superfamily: aldehyde reductase
 Query Match 30.58; Score 527.5; DB 2; Length 309;
 Best Local Similarity 42.3%; Pred. No. 7.9e-34;
 Matches 124; Conservative 51; Mismatches 98; Indels 20; Gaps 8;

Best Local Similarity 38.2%; Pred. No. 1.1e-33;
 Matches 120; Conservative 61; Mismatches 118; Indels 15; Gaps 8;
 QY 8 TLSNGVKIPGVGFTF-ASEGSGKGTYYTAVTTALKTGYRHLDCAWYLYNEGEVGGIRDFL 67
 DB 4 TNSGFKMPIIGLGWRME--KEELRLDIIDAIKIGVRLDCAANYKNEAEVGGALTEAF 61
 QY 68 KENPSVKREDIFVCTKWNHLHRYEDVLMSSIDSLKRLGLDYDMFLVHPVIAAEKNGQG 127
 DB 62 TTG-LVKREDLFITKLMSSDHGH--VIEACKDSLKQLDLYDLFLVHIPATKHTGIG 118
 QY 128 --EPKIGDGVKYLKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKSKKAKV 185
 DB 119 TDSALGDDG--VLIDITITISLETTHDMKELVMSGLVRSIGISNYDVLTTRDCLAYSKI 176
 QY 186 MPHANQIEIHPFLNEELVQYCFKSNIMPVAYSPGSONQVPTTGERVS--ENKTLNEIA 243
 DB 177 KPAVNIETHFYFQDRLSVKFCQKHGICVTAHTPLGGATANAWEFGTVSCLLDDPVLKDVA 236
 QY 244 EKGNTLAQVLIANGRLRGVYVLPKSSNPKRIESNFK--SIELSDADFEALNAVAKGRHP 301
 DB 237 EKYQTVAQIVLRWGIQRTVVIPTKSKPRLLENFQVDFQLSKEDMEVKSME--RNY 294
 QY 302 RFVNMKDTFGYDVM 315
 DB 295 RTHQTAKFWGIELY 308
 RESULT 11
 T02542
 probable alcohol dehydrogenase [imported] - Arabidopsis thaliana
 N;Alternate names: aldehyde dehydrogenase homolog F13M22.26
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
 C;Accession: T02542; G84796
 R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.
 submitted to the EMBL Data Library, June 1998
 A;Description: Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence.
 A;Reference number: Z14677
 A;Accession: T02542
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-294 <ROU>
 A;Cross-references: EMBL:AC004684; NID:G3236234; PID:G3236258
 A;Experimental source: cultivar Columbia
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A;Reference number: A84420; MUID:20083487; PMID:10617197
 A;Accession: G84796
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-294 <STO>
 A;Cross-references: GB:AE002093; NID:G3236258; PIDN:AAC23646.1; GSPDB:GN00139
 C;Genetics:
 A;Gene: F13M22.26; At2g37760
 A;Map position: 2
 A;Introns: 36/3; 53/3; 79/2; 108/3; 200/3
 C;Superfamily: aldehyde reductase
 Query Match 30.2%; Score 523.5; DB 2; Length 294;
 Best Local Similarity 38.3%; Pred. No. 2.2e-33;
 Matches 116; Conservative 61; Mismatches 101; Indels 25; Gaps 7;
 QY 7 FTLSNGVKIPGVGFTF-ASEGSGKGTYYTAVTTALKTGYRHLDCAWYLYNEGEVGGIRDF 66
 DB 8 FELNTGAKLPCVGLGTAM-----VATAEQAKIGYRHIDCAIYNEKEIG-GVLKK 60
 QY 67 LKENPSVKREDIFVCTKWNHLHRYEDVLMSSIDSLKRLGLDYDMFLVHPVIAAEKNGQ 126
 DB 61 LIGDGVKREELFITKLSKWSNHLDPVPALEKTLQDLQIDYDLYLHWPASLKKKE-- 118

Query Match 31.7%; Score 548.5; DB 2; Length 325;
Best Local Similarity 40.5%; Pred. No. 2.8e-35;
Matches 130; Conservative 52; Mismatches 102; Indels 37; Gaps 8;

QY 7 FTLSNGVKIPGVGFGTFASEGSGKETVAVTTALKTGVRHLDCAWYVNEGEVGEGRDF 66
DB 15 FTLENGDKIPISIGLGTWRS--GKDETNAVCAALKAGYRHIDTAHYGNEKEIGSGIRE- 71
QY 67 LKENPSVKREDIFVCTKVNHLHYEDVLSIDSLKRLGLDYVDFLHVWPIA----- 120
DB 72 ----SGVPRTDIWTSKLWCNAHRAGLVPLALEKTLQDLNLEVIDAYLHWPFALLSGPE 127
QY 121 -AENGGGEGKIPGDKGVKVLKLTENPEPTWAMEKIYEDRKARSIGVSNWTIADLEKM 179
DB 128 ELPRNEKE-----LIYEDVP--IETWQAMEELLETKGVYIGISFNFNNEVLDRV 176
QY 180 SKFAKVNPHANQIBIHPFLPNEELVQVCFSGKNIMPVAYSPLGSGNOVPTTG-ERVSSENKT 238
DB 177 LKIAKVRPTTHQMLHFLYLPQTEYLEKHKLIQHVSAYSPLANQNDAYNSDISKLIHKT 236
QY 239 INETAERKGN--TLAQLVLIAGLRGRGVVLPKSNPKRIESNPKSIELSDADFEAINAVA 296
DB 237 LVDIANARGEGITPANTAIASWAVKRGTSVLPKSNVSNRIYSLYPLTDKEMEAINNIG 296
QY 297 KGRHF-----RFVNMKQ 308
DB 297 VVRFSGHGKFAKPFMFVGLQD 317

RESULT 7
S46020
probable aldehyde reductase (EC 1.1.1.21) - Yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBR1127; hypothetical protein YBR149W
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Apr-2002
C:Accession: S46020; S28668
R:Etian, K.D.; Koetter, P.; Rose, M.; Becker, J.; Grey, M.; Li, Z.; Niegemann, E.; Sche
S.; Gruenbein, R.; Hedges, P.; Kiesau, P.; Korol, S.; Krems, B.; Proft, M.; Siegers, K.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S46013
A:Accession: S46020
A:Molecule type: DNA
A:Residues: 1-344 <ENT>
A:Cross-references: EMBL:Z36018; NID:9536473; PIDN:CAA85107.1; PID:G536474; GSPDB:GN0000
A:Experimental source: strain S286C
R:Martinez-Soriano, J.P.; Wong, W.M.; van Ryk, D.I.; Nazar, R.N.
J. Mol. Biol. 217, 629-635, 1991
A:Title: A widely distributed "cat" family of repetitive DNA sequences.
A:Reference number: S28668; MUID:91171289; PMID:2005616
A:Accession: S28668
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-55, 'LQSKLDN', 63-70, '74-344 <MAR>
A:Cross-references: EMBL:M95580; NID:9172583; PIDN:AAA35037.1; PID:G172584
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992
C:Genetics:
A:Gene: SGD:ARAL; MIPS:YBR149W
A:Cross-references: SGD:S0000353
A:Map position: 2R
C:Superfamily: aldehyde reductase
C:Keywords: oxidoreductase

Query Match 31.6%; Score 547.5; DB 1; Length 344;
Best Local Similarity 37.4%; Pred. No. 3.6e-35;
Matches 122; Conservative 58; Mismatches 91; Indels 55; Gaps 8;

QY 7 FTLSNGVKIPGVGFGTFASEGSGKETVAVTTALKTGVRHLDCAWYVNEGEVGEGRDF 66
DB 25 FSLNNGVRIPALGLGTANPHEKLAETKQAKAIKAGYRHIDTAWAYETEPFVGEAIKEL 84
QY 67 LKENPSVKREDIFVCTKVNHLHYEDVLSIDSLKRLGLDYVDFLHVWPIA----- 121
DB 85 L-EDGSKREDLFTITKW-----FVLWDEVDRLSNESLKALGLEVDLLQHWPLCF 136

QY 122 EK-----NGQGEKIPGDKGVKVLKLTENPEPTWAMEKIY-----EDRKA 163
DB 137 EKIKDPKGISGLVTPVDDSGKTWVAADGDL-----ETVKLEKIYLPNDHRV 186
QY 164 RSIQVSNWTIADLEKMSKFAKMPHANOIEIHPFLPNEELVQVCFSGKNIMPVAYSPLGSG 223
DB 187 RAIGVSNFSIYELRLIKECRVPTVQVETHPHLPQMLERKFCFMDHIDILITAYSPLGSH 246
QY 224 N----QVPTTGERSSENKTLNEIAEKGGNTLAQVLIAGLRGRGVVLPKSNPKRIESNF 279
DB 247 GAPNLKIP-----LVKLAEKYNVTGNDLLISVHIRQGITVIVPRSLNPVRISSI 296
QY 280 KSIELSDADFEAINAVAKGRHFRFVN 305
DB 297 EFASLTQDELQELNDFGKYPVRFID 322

RESULT 8
A84599
hypothetical protein At2g21250 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: A84599
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; A
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84599
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-309 <STO>
A:Cross-references: GB:AE002093; NID:94567260; PIDN:AAD23673.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g21250
A:Map position: 2
C:Superfamily: aldehyde reductase

Query Match 30.6%; Score 529.5; DB 2; Length 309;
Best Local Similarity 40.2%; Pred. No. 7.9e-34;
Matches 119; Conservative 54; Mismatches 110; Indels 13; Gaps 7;

QY 8 TILSNGVKIPGVGFGTFASEGSGKETVAVTTALKTGVRHLDCAWYVNEGEVGEGRDFL 67
DB 4 TILSNGKMPVIGLGVWRME--KEGIRDILINAIKIGYRHLDCAADRYNETEVGDALTEAF 61
QY 68 KENPSVKREDIFVCTKVNHLHYEDVLSIDSLKRLGLDYVDFLHVWPIAEKNGQG 127
DB 62 KTG-LVKREDLFTITKLWNSDHGH--VIEACKDSLKQLDYLDLFLVHPFVATKHTGVG 118
QY 128 --EPKIGPDGKVLKLTENPEPTWAMEKIYEDRKARSIGVSNWTIADLEKMSKFAKV 185
DB 119 TTDSALGDDG--VLDIDITISLETWHDVNEKLVSGVRSIGISYDVFLTRDCLAYSKI 176
QY 186 MPHANOIEIHPFLPNEELVQVCFSGKNIMPVAYSPLGSGNOVPTTGERSV--ENKTLNEIA 243
DB 177 KPAVNIETHPYFQRDSLVKFCQKHGICVTAHTPLGGATANAEMFGTVSCLDLDPVLKDV 236
QY 244 BKGGNTLAQVLIAGLRGRGVVLPKSNPKRIESNF--STELSDADFEAINAVAK 297
DB 237 EYKKTVAQVLRWGIQKTVIVPKTKPARLEENFQVDFELSKEDNEVKSMER 292

RESULT 9
T02543
aldehyde dehydrogenase homolog At2g37770 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02543; H84796
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, E.
submitted to the EMBL Data Library, June 1998

aldo-keto reductase (EC 1.-.-.-) YPR1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein D9491.8; protein YDR368w
C;Species: Saccharomyces cerevisiae
C;Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 20-Jun-2000
C;Accession: S61163; S61563; JCS457; PC4473
R;Ding, H.
submitted to the EMBL Data Library, June 1995
A;Description: The sequence of S. cerevisiae cosmid 9481.
A;Reference number: S61159
A;Accession: S61163
A;Molecule type: DNA
A;Residues: 1-312 <DIN>
A;Cross-references: EMBL:U28373; NID:g849184; PIDN:AA64804.1; PID:g849189; MIPS:YDR368w
R;Moses, T.; Juhnke, H.; Sterkel, C.; Zimmermann, F.K.
submitted to the EMBL Data Library, July 1994
A;Description: A new yeast gene with homology to the aldo-keto reductase protein family.
A;Reference number: S61562
A;Accession: S61563
A;Molecule type: DNA
A;Residues: 1-312 <MIO>
A;Cross-references: EMBL:X80642; NID:g1103496; PIDN:CAA5686.1; PID:g1103498
A;Experimental source: strain M5
R;Nakamura, K.; Kondo, S.; Kawai, Y.; Nakajima, N.; Ohno, A.
Biochim. Biotechnol. Biochem. 61, 375-377, 1997
A;Title: Amino acid sequence and characterization of aldo-keto reductase from bakers' yeast
A;Reference number: JCS457; MUID:97212041; PMID:9058981
A;Accession: JCS457
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-312 <NAK>
A;Accession: PC4473
A;Molecule type: protein
A;Residues: 2-36;45-50;68-96, 'X';124-136;178-227;287-293;299-304 <NA2>
C;Genetics:
A;Gene: SGD:YPR1
A;Cross-references: MIPS:YDR368w; SGD:S0002776
A;Map position: 4R
C;Complex: monomer
C;Function:
A;Description: oxidoreductase; NADPH-dependent keto ester reductase; catalyzes the reduction of aldo-keto reductase
C;Superfamily: aldehyde reductase
C;Keywords: monomer; NADP; oxidoreductase
F;2-312/Product: aldo-keto reductase #status experimental <MAT>

Query Match 32.1%; Score 555; DB 2; Length 312;
Best Local Similarity 40.1%; Pred. No. 8.1e-36;
Matches 128; Conservative 60; Mismatches 97; Indels 34; Gaps 9;

QY 2 SNGKFTLSNGVKIPGVGFTFASGSGKGTYYTAVTALKTGVRHLDCAWYVINEGEVGE 61
DB 7 NSSATLKLNTGASIPVLGFTGWSVDNNG--YHSVIAALKAGYRHDAAAIYLNEEVEGR 64

QY 62 GIRDFLKENSVKREDIFVCTKWNHHLRYEDVLWSDSLKRLGLDYVDMFLVHWPIAA 121
DB 65 AIKD-----SGVPREPIITKLNGTEQR--DPEALNKSRLGLDYVDMFLVHWVPL 117

QY 122 EKNQGGEPKIGPDGKYVILKDLTE-----NPEPTWRAMEKIYEDRKARSIGVSNW 171
DB 118 KTDREV-----TDGNVLICPTLEDGTVIDTKENWFIKTMELQELPKTKTAVGVSNF 171

QY 172 TIADLEK--SKFAKVPFHANQIEHPFLPNEELVOYCFSKNIMPVAYSPLGSONQVPTT 229
DB 172 SINNIKELLESNNKVVPAHQIEHLLPQDELIAFCKEKGIVVEAYSPPFGSAN-APLL 230

QY 230 GERVENSKTLNIAEKGGNTLAQVLIAGLRRGVVLPKSSNPKRIESNFKSIELSDAF 289
DB 231 KE-----QAIDNAKHGVEPAQLIISWSIQRGVYVLAKSWNPERIVSNFKIETLPDDF 285

QY 290 EAINAVAKGRHF-RFVNKK 307
DB 286 KTIISLNSKVGHTKRVDMK 304

RESULT 5
S76143
probable aldehyde reductase (EC 1.1.1.-) - Synchocystis sp. (strain PCC 6803)
C;Species: Synchocystis sp.
A;Variety: PCC 6803
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: S76143
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis sp.
S.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S76143
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-327 <KAN>
A;Cross-references: EMBL:D90914; GB:AB001339; NID:gl653477; PIDN:BAAL8402.1; PID:gl65348;
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG
A;Superfamily: aldehyde reductase
C;Keywords: oxidoreductase

Query Match 31.9%; Score 553; DB 1; Length 327;
Best Local Similarity 38.1%; Pred. No. 1.2e-35;
Matches 117; Conservative 69; Mismatches 105; Indels 16; Gaps 6;

QY 5 KFTLSNGVKIPGVGFTFASGSGKGTYYTAVTALKTGVRHLDCAWYVINEGEVSGIR 64
DB 11 KYFLSNGEQIPALGLGFWKS--SPQVVGQAVEQALDGLYRHLDCALYGNIEAIGATLA 68

QY 65 DFLKENPVSKEEDIFVCTKWNHHLRYEDVLWSDSLKRLGLDYVDMFLVHWPIAAEK 124
DB 69 NAFKKG-VKKEELWITSKLSNAHPDPAVLPALKETLQDLGLDYLDLYLHWPIV-- 124

QY 125 CGGEPKIG-PDGKVVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTTADLEKSKFA 183
DB 125 ---QPDVGFPSGQLLEPTFASLEGTWQALEKAVDLGLCHHIGVSNFSLKLEKLVLSMA 181

QY 184 KVMFHANQIEHPFLPNEELVOYCFSKNIMPVAYSPLGSONQVPTTGERVSENKTL---- 239
DB 182 RIPPAVQVVELHPYLOQSDLLTFANSQNILTATVSPGSGDR-PAAFQQAEPKLLTDPV 240

QY 240 -NETAEKGGNTLAQVLIAGLRRGVVLPKSSNPKRIESNFKSIELSDADFEAINAVAKG 298
DB 241 INGLAAEQCSAAQVLLAWAIQRTVTIPKSNPERLEQNLRRAADITLTDSEMAKIALLD 300

QY 299 RHREPVN 305
DB 301 RHRYIVS 307

RESULT 6
T39169
probable oxidoreductase [imported] - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C;Accession: T50378; T39169
R;Connor, R.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21832
A;Accession: T50378
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-325 <CO2>
A;Cross-references: EMBL:AL021815; PIDN:CAA16997.1; GSPDB:GN00067; SPDB:SPBC8E4.04
A;Experimental source: strain 972h-; cosmid c8E4
C;Genetics:
A;Gene: SPDB:SPAC8E4.04
A;Map position: 2
C;Superfamily: aldehyde reductase

A;Map position: 15R
C;Superfamily: aldehyde reductase
C;Keywords: oxidoreductase

Query Match 33.2%; Score 574.5; DB 1; Length 312;
Best Local Similarity 40.8%; Pred. No. 2.4e-37;
Matches 127; Conservative 54; Mismatches 85; Indels 45; Gaps 7;
QY 5 KFTLSNGVPIGVGFGTFASEGSGKETVAVTALKTGVRHLDCAWYVNLNCEGEGIR 64
DB 10 KILSLNTGAQIPQIGLGTQWSK--ENDAYKAVLTALKDGYRHDTAAIYENEDQVQA 67
QY 65 DFLKENSVRKEDI FVCTKWNHLHRYEDVLSIDSLKRLGLDYVDMFLVHPPIAAEK 124
DB 68 D-----SGVPRFEIFVTKLCTQHPEV--ALDQSLKRLGLDYVDMFLVHPPIAA 117
QY 125 GQGEPIKPGDKVILKDLTENP-----PTWAMEKIYEDRKARSIGV 168
DB 118 -----DPAIKNEDILSVPTKDGSRVDITNWNFKTWELQELPKTKTAVGV 168
QY 169 SNWTIADLEK--SKFAKVMPHANOIEHPFLNPEELVQYCFKSNIMPVAYSPLGSONQV 226
DB 169 SNFSINNKLKOLLASQGNKLTPAANQVIEHPPLPQDELINCFKSGIIVEAYSP 224
QY 227 PTTGERSYENKTLNEIAEKGNTLAQVLIAGLRRGVVLPKSSNPKRIESNFKSI 286
DB 225 --TDAPILKEPVEILEIAKNNVQGHVVISWHVQRGVGVVLPKSNPDRINKRIT 282
QY 287 ADPEAINAVAK 297
DB 283 EDPEAINNISK 293

RESULT 2

T38413
A;Description: probable oxidoreductase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T38413
R;Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A;Reference number: 221792
A;Accession: T38413
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-321 <BRO>
A;Cross-references: EMBL:Z73100; PIDN:CAA97364.1; GSPDB:GN00066; SPDB:SPAC2671.07
A;Experimental source: strain 972h-; cosmid c26f1
C;Genetics:
A;Gene: SPDB:SPAC2671.07
A;Map position: 1
C;Superfamily: aldehyde reductase

Query Match 32.8%; Score 568; DB 2; Length 321;
Best Local Similarity 43.0%; Pred. No. 8.2e-37;
Matches 131; Conservative 59; Mismatches 89; Indels 26; Gaps 10;
QY 7 FTLSNGVKIPGVGFGTFASEGSGKETVAVTALKTGVRHLDCAWYVNLNCEGEGIR 66
DB 16 FTLDAGSKI PGLGLGTWRSE--PNQTKNAVKTALQYGRHIDAAIYGNEDVGDGIKE- 72
QY 67 LKENPSVKREDI FVCTKWNHLHRYEDVLSIDSLKRLGLDYVDMFLVHPPIAAEK 126
DB 73 ----SGVPRKDIWTSKLMCNAAPEAVPKALEKTLKDLVLDYLIHWPVSF-KTGE 127
QY 127 GEPKIPGDKGVILKDLTENP-EPTWAMEKIYEDRKARSIGVSNWTIADLEKMSKFAK 195
DB 128 DXFPKDKGNLIYEK-----NPIETWAMEKLETKGVRHIGLSNFDNLEKILKAVK 193
QY 186 MPHANOIEHPFLNPEELVQYCFKSNIMPVAYSPLGSON-----QVPTTGERSYEN 240
DB 184 KPAVHQMELHPFLPQTEFVEKHKGLGHVTAYSFPFGNQNTIYESKIP----KLIEHETIQ 239

QY 241 EIAE-KG-GNTLAQVLIAGLRRGVVLPKSSNPKRIESNFKSIELSDADFEAINAVAK 298
DB 240 KIAKSGEGVTGATIAVGAITRGTSVIPKSVNQRKISNFKYIPLTKEDVDEINSI--G 297
QY 299 RHFRF 303
DB 298 IPARF 302

RESULT 3

S78113
A;Description: aldehyde reductase (NADPH) (EC 1.1.1.-) - fungus (Sporidiobolus salmonicolor)
N;Alternate names: NADPH-dependent aldehyde reductase
C;Species: Sporidiobolus salmonicolor, Sporobolomyces salmonicolor
C;Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 11-Jun-1999
C;Accession: S78113; S23931
R;Kita, K.; Matsuzaki, K.; Hashimoto, T.; Yanase, H.; Kato, N.; Chung, M.C.; Kataoka, M.;
submitted to the EMBL Data Library, May 1995
A;Description: Cloning of the aldehyde reductase gene from a red yeast, Sporobolomyces se
A;Reference number: S78113
A;Accession: S78113
A;Molecule type: DNA
A;Residues: 1-323 <KIT>
A;Cross-references: EMBL:U26463; NID:G1142697; PIDN:AAE17362.1; PID:G1142698
A;Experimental source: strain AKU 4429
R;Kataoka, M.; Sakai, H.; Morikawa, T.; Katoh, M.; Miyoshi, T.; Shimizu, S.; Yamada, H.
Biochim. Biophys. Acta 1122, 57-62, 1992
A;Title: Characterization of aldehyde reductase of Sporobolomyces salmonicolor.
A;Reference number: S23931; MUID:92338224; PMID:1633196
A;Accession: S23931
A;Molecule type: protein
A;Residues: 2-52 <KAT>
A;Experimental source: strain AKU 4429
C;Genetics:
A;Introns: 22/3; 47/3; 204/2; 242/2
C;Function:
A;Description: may be involved in detoxification of reactive carbonyl-containing compounds
A;Note: the authors use EC 1.1.1.2
C;Superfamily: aldehyde reductase
C;Keywords: NADP; oxidoreductase
F;2-323/Product: alcohol dehydrogenase (NADP-) #status experimental <MAT>

Query Match 32.3%; Score 559.5; DB 2; Length 323;
Best Local Similarity 39.3%; Pred. No. 3.8e-36;
Matches 127; Conservative 60; Mismatches 119; Indels 17; Gaps 6;

QY 6 FTLSNGVKIPGVGFGTFASEGSGKETVAVTALKTGVRHLDCAWYVNLNCEGEGIR 65
DB 4 FTTLTNGASLELVGYGTW--QAAPGEVQGVKVAIETGYRHLDLAKVYSNQPEVGAATKE 61
QY 66 FLKENPSVKREDI FVCTKWNHLHRYEDVLSIDSLKRLGLDYVDMFLVHPPIAAEK 125
DB 62 ----AGVKREDLFITSKLWNNSHPRPEQVEPALDDTLKELGLEVLDLYLIHWPVAF 116
QY 126 QGEPIKPGDKGVILK-DLTNPETPTWAMEKIYEDRKARSIGVSNWTIADLEKMSKFAK 184
DB 117 DITQNLFFKANDKEVKLDLEYSLVDTWKAIVKLLDTGKVAIGVSNFADKAVDAI 176
QY 185 VMPHANOIEHPFLNPEELVQYCFKSNIMPVAYSPLGSONQVPTTG-ERSYENKTLNEIA 243
DB 177 VTPSNVQIERHPLLLOPELIAHAKNHIHTAYSPLGN-----TVGAPLIVQHPEIK 232
QY 244 EKGNTLAQVLIAGLRRGVVLPKSSNPKRIESNFKSIELSDADFEAINAVAKGRHFR 303
DB 233 EKNGCTPAQVLIAGLRRGVVLPKSVTPSRIGENFKQVLSQEDVDVAVSKLGE 292
QY 304 VMKDTFG----YDVTPEETAKN 322
DB 293 YNIPCTYSPKMDINVFGEDEKS 315

RESULT 4

S61163

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 14:01:37 ; Search time 20 seconds
(without alignments)
1563.112 Million cell updates/sec

Title: US-10-004-115B-1

Perfect score: 1731

Sequence: 1 MSGNGKFTLLSGVKGIPGVGF.....MKDITFDYDWPETAKNLISA 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	574.5	33.2	312	1 S22846	probable aldehyde
2	568	32.8	321	2 T38413	probable oxidoredu
3	559.5	32.3	323	2 S78113	aldehyde reductase
4	555	32.1	312	2 S61163	aldo-keto reductas
5	553	31.9	327	1 S76143	probable aldehyde
6	548.5	31.7	325	2 T39169	probable oxidoredu
7	547.5	31.6	344	1 S46020	probable aldehyde
8	529.5	30.6	309	2 T02543	hypothetical prote
9	529	30.5	290	2 T02543	aldehyde dehydroge
10	527.5	30.5	309	2 B84599	hypothetical prote
11	523.5	30.2	284	2 T02542	probable alcohol d
12	523.5	30.2	313	2 T09670	absisic acid acti
13	521	30.1	320	2 T48188	aldose reductase-1
14	516	29.8	280	2 C95172	oxidoreductase, al
15	516	29.8	280	2 C98038	conserved hypothet
16	509	29.4	281	2 A37990	aldose reductase h
17	508.5	29.4	281	2 D86658	oxidoreductase ycg
18	503	29.1	276	2 C70040	plant-metabolite d
19	501.5	29.0	310	2 T17013	D-sorbitol-6-phosp
20	500	28.9	280	2 D69988	plant metabolite d
21	499.5	28.9	316	1 A60603	aldehyde reductase
22	491.5	28.4	350	2 B84797	probable alcohol d
23	490.5	28.3	316	2 A59021	aldehyde reductase
24	490	28.3	274	2 AC1535	oxydoreductases ho
25	488.5	28.2	284	1 A32950	probable aldehyde
26	487.5	28.2	294	2 AC3642	2,5-diketo-D-gluc
27	486.5	28.1	316	1 A39763	aldehyde reductase
28	486	28.1	323	1 A56424	estradiol 17beta-d
29	485	28.0	274	2 D97141	probable aldo/keto

30	485	28.0	316	2 A53440	aldose reductase h
31	484.5	28.0	316	2 I49484	aldehyde reductase
32	483	27.9	274	2 AG1177	oxydoreductases ho
33	480	27.7	278	2 A84131	plant-metabolite d
34	478	27.6	315	1 A35452	aldehyde reductase
35	477	27.6	325	2 A33851	alcohol dehydrogen
36	475.5	27.5	323	1 A45366	20alpha-hydroxyste
37	473	27.3	325	1 JN0629	alcohol dehydrogen
38	466	26.9	316	2 T26766	hypothetical prote
39	465.5	26.9	302	2 A34406	aldehyde reductase
40	461	26.6	326	1 S41120	cholestenone 5beta
41	457	26.4	323	2 JC5240	3alpha-hydroxycho
42	456.5	26.4	295	1 S30383	morphine 6-dehydro
43	456	26.3	323	2 I73676	chlorodecone reduct
44	455.5	26.3	329	2 I53872	dihydrodiol dehydr
45	454	26.2	276	2 F83919	plant-metabolite d

ALIGNMENTS

RESULT 1

S22846 probable aldehyde reductase (EC 1.1.1.-) GCY1 - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein O3269; protein YOR120W; protein YOR3269W

C:Species: Saccharomyces cerevisiae

C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: S22846; S60986; S61678; S67005; S63863

R:Oechsner, U.; Magdolen, V.; Bandlow, W.

FEBS Lett. 238, 123-128, 1988

A:Title: A nuclear yeast gene (GCY) encodes a polypeptide with high homology to a vertebr

A:Reference number: S22846; MUID:89005653; PMID:2901985

A:Accession: S22846

A:Molecule type: DNA

A:Residues: 1-312 <OE>

A:Cross-references: ENBL:X13228; NID:g3737; PIDN:CAA31615.1; PID:g3738

R:Wiemann, S.; Rechmann, S.; Benes, V.; Voss, H.; Schwager, C.; Vlcek, C.; Stegemann, J.;

submitted to the ENBL Data Library, August 1995

A:Description: Sequencing of 51 kilobases on the right arm of chromosome XV from S. cerev

A:Reference number: S60983

A:Accession: S60986

A:Molecule type: DNA

A:Residues: 1-312 <WIE>

A:Cross-references: ENBL:X90518; NID:g1050808; PIDN:CAA62107.1; PID:g1050812

R:Benes, V.; Andrade, M.A.; Rechmann, S.; Teodoru, C.; Banreivi, A.; Sander, C.; Valencia,

submitted to the ENBL Data Library, December 1995

A:Description: Nucleotide sequence and analysis of a 130 kb fragment of yeast chromosome

A:Reference number: S61643

A:Accession: S61678

A:Molecule type: DNA

A:Residues: 1-312 <BEN>

A:Cross-references: ENBL:X94335; NID:g1262139; PIDN:CAA64040.1; PID:g1164965

R:Voss, H.; Benes, V.; Rechmann, S.; Teodoru, C.; Schwager, C.; Paces, V.; Ansorge, W.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66965

A:Accession: S67005

A:Molecule type: DNA

A:Residues: 1-312 <VOS>

A:Cross-references: ENBL:Z75028; NID:g1420316; PIDN:CAA99318.1; PID:g1420317; GSPDB:GN000

A:Experimental source: strain S288C

R:Wiemann, S.; Rechmann, S.; Benes, V.; Voss, H.; Schwager, C.; Vlcek, C.; Stegemann, J.;

Yeast 12, 281-288, 1996

A:Title: Sequencing and analysis of 51 kb on the right arm of chromosome XV from Sacchar

A:Reference number: S63860; MUID:97060020; PMID:8904341

A:Accession: S63863

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-312 <WIE>

A:Cross-references: ENBL:X90518; NID:g1050808; PIDN:CAA62107.1; PID:g1050812

A:Note: the nucleotide sequence was submitted to the ENBL Data Library, August 1995

C:Genetics:

A:Gene: SGD:GCY1; MIPS:YOR120W

A:Cross-references: SGD:S0005646; MIPS:YOR120W

Db 85 L-EDGSIKREDLFTTKVM-----PVLWDEVDRLNESLKALGLEVDLLQHWPLCF 136
 QY 122 EK-----NGGEPKIGDGVKILKDLTENPEPTWAMEKIY---EDRKA 163
 Db 137 EKIDPKIGISGLVKTVPDDSGKTWYAADGYL-----ETYKLEKIYLDPNDRHV 186
 QY 164 RSIGVSNWTIADLEKMSKFAKVMPHANQIEIHPFLPNEELVQVCFKSNIMPVAYSPLGSO 223
 Db 187 RAIGVSNFSIEYLERLIKECRVKPTVQVETHPLPQWELRKFVMDILLTAYSPLGSH 246
 QY 224 N-----QVPTTGERVSENKTLNEIAEKGNTLAQVLIANGLRGYYVLPKSNPKRIESNF 279
 Db 247 GAPNLKIP-----LVYKLAEKYNTVGNLLISYHROGTIVIPRSLNVRISSSI 296
 QY 280 KSIELSDADFEAINAVAKGSHFRFVN 305
 Db 297 EFASLTQDELQELNDFGEKYPVRFID 322

RESULT 14
 AAU76345
 ID AAU76345 standard; protein; 344 AA.
 AC AAU76345;
 XX
 DT 21-MAY-2002 (first entry)
 DE Yeast D-arabinose dehydrogenase (ARA).
 KW Ascorbic acid; vitamin C; scurvy; recombinant yeast; enzyme;
 KW L-galactose dehydrogenase; LGDH; L-galactono-1,4-lactone dehydrogenase;
 KW AG; D-arabinose dehydrogenase; ARA; D-arabinono-1,4-lactone oxidase;
 KW AL; L-gulono-1,4-lactone oxidase; GLO; aldono-lactonase; AL.
 XX Saccharomyces cerevisiae.
 XX
 PN WO200210425-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 02-AUG-2001; 2001WO-GB003485.
 XX
 PR 02-AUG-2000; 2000US-00630983.
 XX
 PA (BIOP-) BIPOLO SCARL.
 PA (WHALE/) WHALLEY K.
 XX
 PI Porro D, Sauer M;
 XX
 XX WPI; 2002-217125/27.
 DR N-PSDB; ABK10132.
 XX
 PT Generating ascorbic acid or its salt, involves culturing yeast capable of
 PT converting ascorbic acid precursor into ascorbic acid in medium
 PT comprising ascorbic acid precursor, and isolating ascorbic acid.
 XX
 PS Claim 12; Page 90-91; 95pp; English.
 XX
 CC The invention relates to generating ascorbic acid or its salt, involves
 CC obtaining a recombinant yeast capable of converting an ascorbic acid
 CC precursor into ascorbic acid, culturing the recombinant yeast in a medium
 CC comprising an ascorbic acid precursor, thus forming ascorbic acid, and
 CC isolating the ascorbic acid. Also include are stabilising ascorbic acid
 CC or its salt in a medium, by culturing a yeast in a medium comprising
 CC ascorbic acid or its salt and a recombinant yeast functionally
 CC transformed with a coding region encoding a protein having an enzyme
 CC activity selected from L-galactose dehydrogenase (LGDH), L-galactono-1, 4
 CC -lactone dehydrogenase (AGD), D-arabinose dehydrogenase (ARA), D-
 CC -lactone dehydrogenase (AGD), D-arabinose dehydrogenase (ARA), D-
 CC arabinono-1,4-lactone oxidase (ALO), L-gulono-1,4-lactone oxidase (GLO)
 CC and aldono-lactonase (AL) activity, where the recombinant yeast is capable
 CC of converting to ascorbic acid at least about 25% ascorbic acid precursor
 CC or is capable of producing at least 20 mg ascorbic acid/L medium, when

CC the yeast is cultured in a medium comprising one ascorbic acid precursor.
 CC The ascorbic acid produced (Vitamin C) is a powerful antioxidant, a
 CC deficiency of which causes scurvy in humans. The present sequence
 CC represents yeast ARA
 XX
 SQ Sequence 344 AA;
 Query Match 31.6%; Score 547.5; DB 5; Length 344;
 Best Local Similarity 37.4%; Pred. No. 9.7e-45;
 Matches 122; Conservative 58; Mismatches 91; Indels 55; Gaps 8;
 QY 7 FTLSNGVKIPGVGFTFASGSGKGTETAVTTALKTYRHLDCAWYILNEGEVGEGRDIF 66
 Db 25 FSLNNGVRIPALGVTANPHEKLAETKQAVKAAIKAGYRHIDTAWAYETEFVGEAIKEL 84
 QY 67 LKENPSVKREDIFVCTKVNNHLHRYEDVLM-----SIDDSLKRLGLDYVDMFLVHFI 121
 Db 85 L-EDGSIKREDLFTTKVM-----PVLWDEVDRLNESLKALGLEVDLLQHWPLCF 136
 QY 122 EK-----NGGEPKIGDGVKILKDLTENPEPTWAMEKIY---EDRKA 163
 Db 137 EKIDPKIGISGLVKTVPDDSGKTWYAADGYL-----ETYKLEKIYLDPNDRHV 186
 QY 164 RSIGVSNWTIADLEKMSKFAKVMPHANQIEIHPFLPNEELVQVCFKSNIMPVAYSPLGSO 223
 Db 187 RAIGVSNFSIEYLERLIKECRVKPTVQVETHPLPQWELRKFVMDILLTAYSPLGSH 246
 QY 224 N-----QVPTTGERVSENKTLNEIAEKGNTLAQVLIANGLRGYYVLPKSNPKRIESNF 279
 Db 247 GAPNLKIP-----LVYKLAEKYNTVGNLLISYHROGTIVIPRSLNVRISSSI 296
 QY 280 KSIELSDADFEAINAVAKGSHFRFVN 305
 Db 297 EFASLTQDELQELNDFGEKYPVRFID 322

RESULT 15
 ABP53552
 ID ABP53552 standard; protein; 328 AA.
 AC ABP53552;
 XX
 DT 16-DEC-2002 (first entry)
 DE T. megachiliensis erythrose reductase type I SEQ ID NO:3.
 KW Erythrose reductase; enzyme; erythritol; D-erythrose.
 XX
 OS Trichosporonoides megachiliensis.
 XX
 PN EP1221478-A2.
 XX
 PD 10-JUL-2002.
 XX
 PF 04-JAN-2002; 2002EP-00000321.
 XX
 PR 09-JAN-2001; 2001JP-00001294.
 PR 18-DEC-2001; 2001JP-00384357.
 XX
 XX (NOR) NAT FOOD RES INST MIN AGRIC.
 XX (NIKM) NIKKEN CHEM CO LTD.
 PA
 PA Ookura T, Kasumi T, Asaba E;
 XX
 XX WPI; 2002-659435/71.
 DR N-PSDB; ABQ82227.
 XX
 PT Novel protein having erythrose reductase activity such as erythrose
 PT reductase type I, II or III, useful for production of erythritol on an
 PT industrial scale.
 XX
 XX Claim 17; Page 24-26; 39pp; English.
 XX

XX PN WO200264766-A2.
XX PD 22-AUG-2002.
XX PF 21-DEC-2001; 2001WO-EP015398.
XX PR 22-DEC-2000; 2000EP-00870318.
XX PR 04-JAN-2001; 2001EP-00870002.
XX PR 09-JAN-2001; 2001EP-00870003.
XX PA (JANC) JANSSEN PHARM NV.
XX PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
XX DR WPI; 2002-667002/71.
XX DR N-PSDB; ABQ76569.
XX PT New isolated nucleic acid representing a synthetic BAX-gene, useful as
XX PT medicament for treating, preventing and/or alleviating yeast or fungal
XX PT infections or proliferative disorders, or for preventing apoptosis in
XX PT certain diseases.
XX PS Claim 36; Fig 2; 344pp; English.
XX CC This invention describes a novel nucleic acid representing a synthetic
XX CC Bax gene. The Bax gene of the invention is useful for identifying Bax-
XX CC resistant yeast or fungi, identifying, or obtaining and identifying
XX CC Candida spp. sequences that are differentially expressed in a pathway
XX CC eventually leading to programmed cell death or identifying inhibitors or
XX CC inhibitor sequences of Bax-induced cell death. The products of the
XX CC invention have cytostatic, fungicide; immunosuppressive, virucide and
XX CC vasotropic activity and can be used in vaccines or for gene therapy. The
XX CC isolated nucleic acids, polypeptides, pharmaceutical compositions,
XX CC antisense molecules and antibodies are useful as medicaments or in
XX CC preparing a medicament for treating, preventing and/or alleviating
XX CC diseases associated with yeast or fungi or proliferative disorders, such
XX CC as cancer, or for preventing apoptosis in certain diseases. The compounds
XX CC or polypeptides, or the genetically modified organism are useful for
XX CC preparing a medicament for modifying the endogenous flora of humans and
XX CC other mammals. The vaccine is useful for immunising against yeast or
XX CC fungal infections. Apoptosis-related diseases include autoimmune disease,
XX CC ischaemia, diseases related with viral infections or neurodegenerations.
XX CC This sequence represents a polypeptide associated with the Bax gene
XX CC described in the disclosure of the invention
XX SQ Sequence 295 AA;
Query Match 31.8%; Score 550; DB 5; Length 295;
Best Local Similarity 41.9%; Pred. No. 4.4e-45;
Matches 125; Conservative 56; Mismatches 87; Indels 30; Gaps 8;
QY 3 NGKFTLSNGVKIPGVGFTFASGSGKGTAVTTALTGTGYRHLDCAWYLNVEGVGEG 62
Db 8 NTDFTLNGNKIPAVGLGTWQAT-NEDEAYRAVLAKNGYKHIDTAIYGNVEGVGKA 66
QY 63 IRDFLKNPSYKREDIFVCTKVMNHLHRYEDVLMSIDDSLKRLGLDYVDFVHPVPIAAE 122
Db 67 IKD-----SGVPRSELFVTTKLWNADEK--NIEEALETSLKGLCNVDLYLIHWPSAD 119
QY 123 KNGGGEKIPGDKYVILKOLTENPPTWRAMEKIYED-RKARSIGVSNWTIADLEKM-- 179
Db 120 KSTN---KPYTDFDYV-----DTRYGLQVYKNSKKIRAGVSNFTKKLRLLS 166
QY 180 SKFAKVMFANQIETHPLFNEELVQYCFGKNIHFVAYSPIGSONQVPTTGCVSVENKTL 239
Db 167 SEGVDVVPVANNQIEAHPLLTQPELYDLKKEGIVLEAYSPIGS-----INSLFKNETI 220
QY 240 NEIAEKGNTLAQVLIANGRLGCVVLPKSNPNKRIEISNFKSIELSDADFEANAVAK 297
Db 221 VKIAEKNGVEPAQVLVSNAIQKRWLPKSVTESRSVSNLKTFTLPSEDEFETLNKUSE 278

RESULT 13
ABC93050
ID ABG93050 standard; protein; 344 AA.
XX AC ABG93050;
XX DT 21-NOV-2002 (first entry)
XX DE S. cerevisiae BAX-associated protein fragment SEQ ID 58.
XX KW Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
XX KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
XX KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
XX KW neurodegeneration; cell death.
XX OS Saccharomyces cerevisiae.
XX PN WO200264766-A2.
XX PD 22-AUG-2002.
XX PF 21-DEC-2001; 2001WO-EP015398.
XX PR 22-DEC-2000; 2000EP-00870318.
XX PR 04-JAN-2001; 2001EP-00870002.
XX PR 09-JAN-2001; 2001EP-00870003.
XX PA (JANC) JANSSEN PHARM NV.
XX PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
XX DR WPI; 2002-667002/71.
XX DR N-PSDB; ABQ76316.
XX PT New isolated nucleic acid representing a synthetic BAX-gene, useful as
XX PT medicament for treating, preventing and/or alleviating yeast or fungal
XX PT infections or proliferative disorders, or for preventing apoptosis in
XX PT certain diseases.
XX PS Claim 36; Fig 1; 344pp; English.
XX CC This invention describes a novel nucleic acid representing a synthetic
XX CC Bax gene. The Bax gene of the invention is useful for identifying Bax-
XX CC resistant yeast or fungi, identifying, or obtaining and identifying
XX CC Candida spp. sequences that are differentially expressed in a pathway
XX CC eventually leading to programmed cell death or identifying inhibitors or
XX CC inhibitor sequences of Bax-induced cell death. The products of the
XX CC invention have cytostatic, fungicide; immunosuppressive, virucide and
XX CC vasotropic activity and can be used in vaccines or for gene therapy. The
XX CC isolated nucleic acids, polypeptides, pharmaceutical compositions,
XX CC antisense molecules and antibodies are useful as medicaments or in
XX CC preparing a medicament for treating, preventing and/or alleviating
XX CC diseases associated with yeast or fungi or proliferative disorders, such
XX CC as cancer, or for preventing apoptosis in certain diseases. The compounds
XX CC or polypeptides, or the genetically modified organism are useful for
XX CC preparing a medicament for modifying the endogenous flora of humans and
XX CC other mammals. The vaccine is useful for immunising against yeast or
XX CC fungal infections. Apoptosis-related diseases include autoimmune disease,
XX CC ischaemia, diseases related with viral infections or neurodegenerations.
XX CC This sequence represents a polypeptide associated with the Bax gene
XX CC described in the disclosure of the invention
XX SQ Sequence 344 AA;
Query Match 31.6%; Score 547.5; DB 5; Length 344;
Best Local Similarity 37.4%; Pred. No. 9.7e-45;
Matches 122; Conservative 58; Mismatches 91; Indels 55; Gaps 8;
QY 7 FTLSNGVKIPGVGFTFASGSGKGTAVTTALTGTGYRHLDCAWYLNVEGVGEGIDF 66
Db 25 FSLNNGVRIPALGLGTANPHEKLAETKQAKVAAIKAGYRHDITAWAYETFPVGEAIKEL 84
QY 67 LKENPSVKREDIFVCTKVMNHLHRYEDVLW-----SIDDSLKRLGLDYVDFVHPVPIAA 121

PT Carbonyl reductase gene and protein - used in the production of R-gamma-
 PT substituted-beta-hydroxybutyrate, for use in drugs and agrochemicals.
 XX PS Claim 1; Page 9-10; 14pp; Japanese.
 XX The present sequence is that of a carbonyl reductase (ALD) isolated from
 CC Sporobolomyces salmonicolor IF01038. The enzyme is used in the synthesis
 CC of an (R)-gamma-substituted-beta-hydroxybutyrate (GSBH) from a gamma-
 CC substituted acetoacetate. The GSBH is useful as a raw material for the
 CC synthesis of drugs and agricultural chemicals. (Updated on 16-OCT-2003 to
 CC standardise OS field)
 XX SQ Sequence 323 AA;

Query Match 32.3%; Score 559.5; DB 2; Length 323;
 Best Local Similarity 39.3%; Pred. No. 5.8e-46;
 Matches 127; Conservative 60; Mismatches 119; Indels 17; Gaps 6;
 QY 6 TFLTSLNGVKIPGVGFTFASGSGKGETYTAVTTALKTGYRHLDCAWYLNNEGVEGEGIRD 65
 DB 4 TTTILNTGASLELVGYGTW--QAAPGEVGGQGVKVAIETGYRHLDLAKVYSNQPEVGAAIKE 61
 QY 66 FLKPNPSVKREDIFVCTKQVNHHLRYEDVLWSDSLKRLGLDYVDMFLVHWPPIAAEKNG 125
 DB 62 -----AGVKREDLFTSKLWNSHRPEQVEPALDDTLKELGLDYLDLYLHWPVAFPPFG 116
 QY 126 QGEPKIGPDGKYVLK-DLTENPEPTWRAMEKIYEDRKARSIGVSNWTTIADLEKMSKFAK 184
 DB 117 DITQNLFPKANDKEVKLDLESLVDTWKAMVLLDTCKYKAIGVSNFDAQWDAIIETNG 176
 QY 185 VMPHANOIETHPFLPNEELVOYCFSKNIMPVAYSPILGSONQVPTTG-ERVSENKTLNEIA 243
 DB 177 VTFSVNOIERHPLLOPELIAHAKAKNIHITAYSPLGN-----TVGAPLLVQHPFKRIA 232
 QY 244 EKGNTLAQVLIAGLRGVVLPKSNPKRIESNFKSIELSADPEAINAVAKGRHFRF 303
 DB 233 EKGCTPAQVLIAGVGGSHVIPKSVTPSRIGENFKQVLSQEDVDVSKLGEVGGRRR 292
 QY 304 VNMKDTFG-----YDVWPEETAKN 322
 DB 293 YNIPCTYSPKWDINVFGEDEKS 315

RESULT 11
 ID AAG63564
 XX AAG63564 standard; protein; 313 AA.

AC AAG63564;

DT 15-OCT-2001 (first entry)

DE Amino acid sequence of a ketone reductase.

XX Methionine gamma-lyase; mdeA gene; free folding energy; gene shuffling;
 KW directed evolution; molecular breeding; ketone reductase.

XX Saccharomyces cerevisiae.

XX WO20015342-A2.

XX 02-AUG-2001.

XX 31-JAN-2001; 2001WO-US003186.

XX 31-JAN-2000; 2000US-00494921.

XX 08-DEC-2000; 2000US-00734237.

XX (BIOC-) BIOCATALYTICS INC.

XX Rozzell DJ, Bui P, Hua L;

XX WPI; 2001-483235/52.

XX N-PSDE; AAH74581.

XX PT Designing synthetic nucleic acid sequences for improved amplification,
 PT expression in host cell, by comparing free energy of folding of a
 PT starting polynucleotide and a modified polynucleotide having a codon
 PT replacement.
 XX Claim 6; Page 69-70; 117pp; English.
 XX The present sequence represents a ketone reductase. The polynucleotide
 CC sequence was modified using the method of the invention. The
 CC specification describes a method for designing a synthetic
 CC polynucleotide. The method comprises providing a starting polynucleotide,
 CC determining the predicted free energy of folding per base of the
 CC polynucleotide, modifying the polynucleotide by replacing a codon with a
 CC different codon to provide a modified polynucleotide, determining free
 CC energy of folding per base of the modified polynucleotide, and comparing
 CC this with that of the original polynucleotide. The method is useful for
 CC developing nucleic acid sequences that enhance expression of the
 CC protein in a heterologous host. The design and preparation of the
 CC synthetic genes are used in application of gene shuffling, directed
 CC evolution and molecular breeding methods. The method allows expression of
 CC genes from various organisms such as mammals, plants, fungi and
 CC bacteria in prokaryotic hosts, such as Escherichia coli and eukaryotic
 CC hosts at commercially viable levels, in particular proteins with low
 CC yield such as methionine gamma-lyase from *P. putida*

XX SQ Sequence 313 AA;

Query Match 32.1%; Score 555; DB 4; Length 313;
 Best Local Similarity 40.1%; Pred. No. 1.5e-45;

Matches 128; Conservative 60; Mismatches 97; Indels 34; Gaps 9;

QY 2 SNGKFTTSLNGVKIPGVGFTFASGSGKGETYTAVTTALKTGYRHLDCAWYLNNEGVEGE 61
 DB 8 NSSATLKLNTGASIPVLGFGTWRSDNNG--YHSVIAALKAGYRHIDAAAIYLNNEEYVGR 65

QY 62 GIREPLKENPSVKREDIFVCTKQVNHHLRYEDVLWSDSLKRLGLDYVDMFLVHWPPIAA 121
 DB 66 AIKD-----SGVPREEIPITTKLWGTQOR--DPAALNKLKRLGLDYVDMFLVHWPVPL 118

QY 122 EKNQGEPEKIGPDGKYVLKDLTE-----NPEPTWRAMEKIYEDRKARSIGVSNW 171
 DB 119 KTRV-----TDGNVLCTPLEDGTVDIDTKENFIKTWELMQELPKTKTKAVGVSNF 172

QY 172 TIADLEKM--SKPAKVMPHANOIETHPFLPNEELVOYCFSKNIMPVAYSPILGSONQVPTT 229
 DB 173 SINNIKELLESFNKKVWPATNOIEIHPLLPQDELIAFCKEKGIVVEAYSPFGSAN-APLL 231

QY 230 GERVSENKTLNEIAEKGGNTLAQVLIAGLRGVVLPKSNPKRIESNFKSIELSADPE 289
 DB 232 KE-----QAIDMAKKGVEPAQLIISWISIQGVVLAKSVNPERIVSNFKIFTLPEDDF 286

QY 290 BAINAVAKGRHF-RPVNMK 307

DB 287 KTISNLSKVHGTGRVDMK 305

RESULT 12

ABG93303

ID ABG93303 standard; protein; 295 AA.

XX AC ABG93303;

XX 21-NOV-2002 (first entry)

XX C. albicans BAX-associated protein fragment SEQ ID 564.

XX Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
 KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
 KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
 KW neurodegeneration; cell death.

OS Candida albicans.

XX SQ Sequence 312 AA;
Query Match 33.1%; Score 573.5; DB 2; Length 312;
Best Local Similarity 40.8%; Pred No. 2.3e-47;
Matches 127; Conservative 54; Mismatches 85; Indels 45; Gaps 7;
QY 5 KTFILSNKIPGVGFTFASGSGKETYTAVTTALKTGYRHLDCAWYLNAGEVGEIGR 64
DB 10 KILSLNTGAQIPQIGLGTWQSK--ENDAYKAVLTALKDGYRHDTAAIYENEDQVQAIAK 67
QY 65 DFLKENSVKREDIFVCTKWNHHLRYEDVLWSDDSLKRLGLDLYDMFLVHWPPIAAEKN 124
DB 68 D-----SGVPREIEFVTKLWCTQHHEPEV--ALDQSLKRLGLDLYDMFLVHWPPIAAEKN 117
QY 125 GQGEPKIPGDPGVKLVLDLTENPE-----PTWRAMEKIYEDRKARSIGV 168
DB 118 -----DPAYIKNEDILSVPTKDGSRVADITNNFIKTWELMQELPKTKTKAVGV 168
QY 169 SNWTIADLEKM--SKFAKVMPHANOIEIHPFLPNEELVQYCFKSNIMPVAYSPLGSONQV 226
DB 169 SNFSINNKLKLLASQGNKLTAAANQVEIHPFLPQDELINFCCKGIVVEAYSPLGS---- 224
QY 227 PTTGERVSENKTLNEIAEKGGNTLAQVLIANGRLRGYVVLPKSNPKRIEKNFKSIELSD 286
DB 225 --TDAPLLKEPVLIEIAKKNVQGHVVISHWVQGVVVLPKSVNPDRIKTKRKTFTLST 282
RESULT 9
AAW29219
ID AAW29219 standard; protein; 312 AA.
XX AAW29219;
XX
XX
XX 02-MAR-1998 (first entry)
XX
XX S. cerevisiae uronate dehydrogenase variant Y47E.
XX uronate reductase; uronate dehydrogenase; Saccharomyces cerevisiae;
XX synthesis; L-ascorbic acid; variant.
XX Saccharomyces cerevisiae.
XX
XX Key Location/Qualifiers
XX Misc-difference 47
XX /label= Y57E
XX /note= "wild type tyrosine replaced by glutamine"
XX DE19604798-A1.
XX
XX 14-AUG-1997.
XX
XX 09-FEB-1996; 96DE-01004798.
XX
XX 09-FEB-1996; 96DE-01004798.
XX
XX (HERB-) HERBSTREITH & FOX PEKTIN-FAB NEUENEUEERG.
XX
XX Mattes R, Kulbe K;
XX WPI; 1997-403701/38.
XX
XX Recombinant uronate reductase and L-gulonono-gamma-lactone oxidase - used
XX in L-ascorbic acid production from D-glucuronic or galacturonic acids.
XX
XX Claim 31; Page; 26pp; German.
XX
XX The present sequence represents a uronate dehydrogenase (reductase) Y47E
XX variant. The variant can be used for synthesis of L-ascorbic acid. DNA

CC encoding L-gulonono-gamma-lactone oxidase (GluOx) is also used in the
CC production of L-ascorbic acid. The uronate dehydrogenase has higher
CC affinity for substrate (D-glucuronic or galacturonic acids) and coenzymes
CC than the similar enzyme from lipomyces starkeyi, also better stability
CC and greater yield when produced recombinantly. NB. This sequence was
CC created using the wild type uronate dehydrogenase sequence given in
CC Figure 3 of the specification
XX
XX SQ Sequence 312 AA;
Query Match 32.7%; Score 565.5; DB 2; Length 312;
Best Local Similarity 40.5%; Pred No. 1.4e-46;
Matches 126; Conservative 54; Mismatches 86; Indels 45; Gaps 7;
QY 5 KTFILSNKIPGVGFTFASGSGKETYTAVTTALKTGYRHLDCAWYLNAGEVGEIGR 64
DB 10 KILSLNTGAQIPQIGLGTWQSK--ENDAYKAVLTALKDGERHDTAAIYRNEDQVQAIAK 67
QY 65 DFLKENSVKREDIFVCTKWNHHLRYEDVLWSDDSLKRLGLDLYDMFLVHWPPIAAEKN 124
DB 68 D-----SGVPREIEFVTKLWCTQHHEPEV--ALDQSLKRLGLDLYDMFLVHWPPIAAEKN 117
QY 125 GQGEPKIPGDPGVKLVLDLTENPE-----PTWRAMEKIYEDRKARSIGV 168
DB 118 -----DPAYIKNEDILSVPTKDGSRVADITNNFIKTWELMQELPKTKTKAVGV 168
QY 169 SNWTIADLEKM--SKFAKVMPHANOIEIHPFLPNEELVQYCFKSNIMPVAYSPLGSONQV 226
DB 169 SNFSINNKLKLLASQGNKLTAAANQVEIHPFLPQDELINFCCKGIVVEAYSPLGS---- 224
QY 227 PTTGERVSENKTLNEIAEKGGNTLAQVLIANGRLRGYVVLPKSNPKRIEKNFKSIELSD 286
DB 225 --TDAPLLKEPVLIEIAKKNVQGHVVISHWVQGVVVLPKSVNPDRIKTKRKTFTLST 282
RESULT 10
AAW296294
ID AAR96294 standard; protein; 323 AA.
XX AAR96294;
XX
XX 16-OCT-2003 (revised)
DT 18-NOV-1996 (first entry)
XX
XX Carbonyl reductase.
XX
XX carbonyl reductase; R-gamma-substituted-beta-hydroxybutyrate;
XX gamma-substituted acetoacetate; raw material; drug synthesis;
XX agricultural chemical production.
XX
XX Sporidiobolus salmonicolor; IF01038.
XX
XX Key Location/Qualifiers
XX FT Protein 2..323
XX FT /label= mature_protein
XX
XX JF08103269-A.
XX
XX 23-APR-1996.
XX
XX 07-OCT-1994; 94JP-00244090.
XX
XX 07-OCT-1994; 94JP-00244090.
XX
XX (ELED) DENKI KAGAKU KOGYO KK.
XX
XX WPI; 1996-253873/26.
XX
XX N-PSDB; AAT29159.
XX

Qy	5	KYFTLSNGVKIPGVCGFTFASGSKGETYTAVTTALKTGYRHLDCAWYILNEGEVCEGIR	64
Db	11	KILSLNTGAQIPQIGLGTSQSK--ENDAYKAVLTALXGYYRHIDTAAYRNEDQVGQAIK	68
Qy	65	DFLKENPSVGEEDIEVCTKWNHLLHRYEDVLWSIDSLKRLGLDYVDMFVHWPIAAEKN	124
Db	69	D-----SGVPREEIFVTKLCTQHHEPEV--ALDOSLKRGLDYVDLYLHWPARL---	118
Qy	125	GOCEPKIGDGKVKILKDLTENPE-----PTWRAMEKIYEDRKASIGV	168
Db	119	-----DPAYIKNEDILSVPTKDGSRADVITNNWNFIKTWELMOELPKTKTKAVGV	169
Qy	169	SNWTTADLEKM--SKFAKWMHPANQIRTHPLPNEELVQVCFSKNIMPVAYSPLGSGNOV	226
Db	170	SNFSINNKLDLASGNGKLTFAAQVEILHPLLPDELFNFKCSKGIYVEAYSPLGS----	225
Qy	227	PTTGERVSENKTLNIAEKGNTLAQVLIAMGLRGVGVLPFKPSNPKRIEISNFKSIELSD	286
Db	226	--TDAPLLKEPVILEIAKNNVQPGHVVISWHVQGVVLPKSVNPDRIKTRKIETLST	283
Qy	287	ADFEAINAVAK	297
Db	284	EDFEAINNISK	294
RESULT 8			
AAW29218			
ID	AAW29218 standard; protein; 312 AA.		
XX	AAW29218;		
DT	02-MAR-1998 (first entry)		
DB	S. cerevisiae uronate dehydrogenase variant R57E.		
XX	uronate reductase; uronate dehydrogenase; Saccharomyces cerevisiae;		
KW	synthesis; L-ascorbic acid; variant.		
XX	Saccharomyces cerevisiae.		
XX	Key	Location/Qualifiers	
FH	Misc-difference 57	/label= R57E	
FT	/note= "wild type arginine replaced by glutamine"		
FT	DE19604798-A1.		
PN	14-AUG-1997.		
PD			
XX	09-FEB-1996; 96DE-01004798.		
PF	09-FEB-1996; 96DE-01004798.		
XX	(HERB-) HERBSTREITH & FOX PEKTIN-PAB NEUENBUERG.		
PA	Mattes R, Kulbe K;		
XX	WPI; 1997-403701/38.		
PI			
DR	Recombinant uronate reductase and L-gulonono-gamma-lactone oxidase - used		
XX	in L-ascorbic acid production from D-glucuronic or galacturonic acids.		
PT	Claim 31; Page; 36pp; German.		
XX			
XX	The present sequence represents a uronate dehydrogenase (reductase) R57E		
CC	variant. The variant can be used for synthesis of L-ascorbic acid. DNA		
CC	encoding L-gulonono-gamma-lactone oxidase (GluOx) is also used in the		
CC	production of L-ascorbic acid. The uronate dehydrogenase has higher		
CC	affinity for substrate (D-glucuronic or galacturonic acids) and coenzymes		
CC	than the similar enzyme from lipomyces starkeyi, also better stability		
CC	and greater yield when produced recombinantly. NB. this sequence was		
CC	created using the wild type uronate dehydrogenase sequence given in		
CC	Figure 3 of the specification		

PA (BIOC-) BIOCATALYTICS INC.
XX Roszell DU, Bui P, Hua L;
XX WPI; 2001-483235/52.
DR N-PSDB; AAH74583.
XX
XX Designing synthetic nucleic acid sequences for improved amplification,
PT expression in host cell, by comparing free energy of folding of a
PT starting polynucleotide and a modified polynucleotide having a codon
XX replacement.
XX
PS Claim 6; Page 72-73; 117pp; English.
XX
XX The present sequence represents a yeast protein. The polynucleotide
CC sequence was modified using the method of the invention. The
CC specification describes a method for designing a synthetic
CC polynucleotide. The method comprises providing a starting polynucleotide,
CC determining the predicted free energy of folding per base of the
CC polynucleotide, modifying the polynucleotide by replacing a codon with a
CC different codon to provide a modified polynucleotide, determining free
CC energy of folding per base of the modified polynucleotide, and comparing
CC this with that of the original polynucleotide. The method is useful for
CC developing nucleic acid sequences that enhance expression of the encoded
CC protein in a heterologous host. The design and preparation of the
CC synthetic genes are used in application of gene shuffling, directed
CC evolution and molecular breeding methods. The method allows expression of
CC genes from various organisms such as mammals, plants, yeast, fungi and
CC bacteria in prokaryotic hosts, such as *Escherichia coli* and eukaryotic
CC hosts at commercially viable levels, in particular proteins with low
CC yield such as methionine gamma-lyase from *P. putida*
XX
SQ Sequence 312 AA;

Query Match 33.2%; Score 574.5; DB 4; Length 312;
Best Local Similarity 40.8%; Pred. No. 1.9e-47;
Matches 127; Conservative 54; Mismatches 85; Indels 45; Gaps 7;

QY 5 KTFILNSGVKIPGVGFTFASGSGKGETYATVTTALKTGYRHLDCAWYILNEGEVGEIGR 64
DB 10 KILSLNTGAQIPQIGLGTWQSK--ENDAYKAVLTALKDGYRHIDTAIYRNEQVQGAIK 67

QY 65 DFLKENPSVKREDIPVCTKVNHLHRYEDVLSWSDSLKRLGLDYDMFLVHWPFAEKN 124
DB 68 D-----SGVPREEIFVTITKLWCTQHHEPEV--ALDQSLKRLGLDYDMFLVHWPFAEKN 117

QY 125 GQGEPIKPGDGKGVILKDLTENPE-----PTWRAMEKIVEDRKARSIGV 168
DB 118 -----DPAYIKNEDILSVPTKDGSRVADITNNFIKTWELMQELPKTKTKAVGV 168

QY 169 SNWTIADLEKM--SKFAKVNPHANQIEIHPFLPNEELVQYCFSKNIMPVAYSPLGSONQV 226
DB 169 SNFSINNKLKDLASQGNKLTAAANQVEIHPFLPQDELINFCCKSGIIVVEAYSPLGS---- 224

QY 227 PTTGERSVSEKTLNEIAEKGNTLAQVLIAGLRGVVVLPSKSNPKRIENFSKIELSD 286
DB 225 --TDAPLKPEVILEIAKNNVQCHVVISHWQVGVVLPKSVNPDRIKINRKITLST 282

QY 287 ADFAINAVAK 297
DB 283 EDFAINNISK 293

RESULT 6
ID ABG93198 standard; protein; 312 AA.
XX AC ABG93198;
XX DT 21-NOV-2002 (first entry)
XX DE S. cerevisiae BAX-associated protein fragment SEQ ID 354.
XX

KW Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
KW neurodegeneration; cell death.
XX
OS Saccharomycetes cerevisiae.
XX
FN W0200264766-A2.
XX
PD 22-AUG-2002.
XX
XX 21-DEC-2001; 2001WO-EP015398.
XX
XX 22-DEC-2000; 2000EP-00870318.
PR 04-JAN-2001; 2001EP-00870002.
PR 09-JAN-2001; 2001EP-00870003.
XX
PA (JANC) JANSSEN PHARM NV.
XX
PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
XX
DR WPI; 2002-667002/71.
DR N-PSDB; ABQ76464.
XX
XX New isolated nucleic acid representing a synthetic BAX-gene, useful as
PT medicament for treating, preventing and/or alleviating yeast or fungal
PT infections or proliferative disorders, or for preventing apoptosis in
PT certain diseases.
XX
PS Claim 36; Fig 1; 344pp; English.
XX
CC This invention describes a novel nucleic acid representing a synthetic
CC Bax gene. The Bax gene of the invention is useful for identifying Bax-
CC resistant yeast or fungi, identifying, or obtaining and identifying
CC *Candida* spp. sequences that are differentially expressed in a pathway
CC eventually leading to programmed cell death or identifying inhibitors or
CC inhibitor sequences of Bax-induced cell death. The products of the
CC invention have cytostatic, fungicide, immunosuppressive, virucide and
CC vasotropic activity and can be used in vaccines or for gene therapy. The
CC isolated nucleic acids, polypeptides, pharmaceutical compositions,
CC antisense molecules and antibodies are useful as medicaments or in
CC preparing a medicament for treating, preventing and/or alleviating
CC diseases associated with yeast or fungi or proliferative disorders, such
CC as cancer, or for preventing apoptosis in certain diseases. The compounds
CC or polypeptides, or the genetically modified organism are useful for
CC preparing a medicament for modifying the endogenous flora of humans and
CC other mammals. The vaccine is useful for immunising against yeast or
CC fungal infections. Apoptosis-related diseases include autoimmune disease,
CC ischaemia, diseases related with viral infections or neurodegenerations.
CC This sequence represents a polypeptide associated with the Bax gene
CC described in the disclosure of the invention
XX
SQ Sequence 312 AA;

Query Match 33.2%; Score 574.5; DB 5; Length 312;
Best Local Similarity 40.8%; Pred. No. 1.9e-47;
Matches 127; Conservative 54; Mismatches 85; Indels 45; Gaps 7;

QY 5 KTFILNSGVKIPGVGFTFASGSGKGETYATVTTALKTGYRHLDCAWYILNEGEVGEIGR 64
DB 10 KILSLNTGAQIPQIGLGTWQSK--ENDAYKAVLTALKDGYRHIDTAIYRNEQVQGAIK 67

QY 65 DFLKENPSVKREDIPVCTKVNHLHRYEDVLSWSDSLKRLGLDYDMFLVHWPFAEKN 124
DB 68 D-----SGVPREEIFVTITKLWCTQHHEPEV--ALDQSLKRLGLDYDMFLVHWPFAEKN 117

QY 125 GQGEPIKPGDGKGVILKDLTENPE-----PTWRAMEKIVEDRKARSIGV 168
DB 118 -----DPAYIKNEDILSVPTKDGSRVADITNNFIKTWELMQELPKTKTKAVGV 168

QY 169 SNWTIADLEKM--SKFAKVNPHANQIEIHPFLPNEELVQYCFSKNIMPVAYSPLGSONQV 226
DB 169 SNFSINNKLKDLASQGNKLTAAANQVEIHPFLPQDELINFCCKSGIIVVEAYSPLGS---- 224

DR WPI; 1997-403701/38.
 XX
 PT Recombinant uronate reductase and L-gulono-gamma-lactone oxidase - used
 PT in L-ascorbic acid production from D-glucuronic or galacturonic acids.
 XX
 XX Claim 31; Page; 26pp; German.
 XX
 CC The present sequence represents a uronate dehydrogenase (reductase) Q29G
 CC variant. The variant can be used for synthesis of L-ascorbic acid. DNA
 CC encoding L-gulono-gamma-lactone oxidase (GluOx) is also used in the
 CC production of L-ascorbic acid. The uronate dehydrogenase has higher
 CC affinity for substrate (D-glucuronic or galacturonic acids) and coenzymes
 CC than the similar enzyme from *Lipomyces starkeyi*, also better stability
 CC and greater yield when produced recombinantly. NB. This sequence was
 CC created using the wild type uronate dehydrogenase sequence given in
 CC Figure 3 of the specification
 XX
 SQ Sequence 312 AA;

Query Match 33.2%; Score 575.5; DB 2; Length 312;
 Best Local Similarity 40.8%; Pred. No. 1.9e-47;
 Matches 127; Conservative 54; Mismatches 85; Indels 45; Gaps 7;
 QY 5 KTFTLSNGVKIPGVGFTFASGSKGTYTAVTTALKTYRHLDCAWYILNEGEVGEIGIR 64
 Db 10 KILSLNTGAQIPQICGLGTWQSK--ENDAYKAVLTALKDGYRHIDTAIYRNEDQVQAIAK 67
 QY 65 DFLKENSVPKREDIFVCTKWNHLHRYEDVLSIDSLKRLGLDYDMFLVHWPAAEKN 124
 Db 68 D-----SGVPREEIFVTIKLWCTQHHEPEV--ALDQSLKRLGLDYDLYLMHWPAAEL--- 117
 QY 125 GQGEPKIGPDKGVILKDLTENPE-----PTWRAMEKIYEDRKARSIGV 168
 Db 118 -----DPAYIKNEDILSVPTKDGSRADVITNNFNKTMQLMQLPCKTKTAVGV 168
 QY 169 SNWTIADLEKM--SKFAKVMPHANQIEIHPPLNEELVOYCFCKNIMPVAYSPLGSONQV 226
 Db 169 SNFSINNKLKOLLASQGNKLTAAQVVEIHPLLPQDELINFCCKSGIWEAYSPLGS---- 224
 QY 227 PTTGERSVENKTLNIEAEKGNNTLAQVLIANGLRGYYVLPKSNPKRIESNFKSIELSD 286
 Db 225 --TDAPLKEPVEILEIAKNNVQPGHVVISWHVQRGYVLPKSNPNDRIKTRKIFTLTST 282
 QY 287 ADPEAINAVAK 297
 Db 283 EDFAINNISK 293

RESULT 4
 AAW29217
 ID AAW29217 standard; protein; 312 AA.
 AC AAW29217;
 XX
 DT 02-MAR-1998 (first entry)
 XX
 DE S. cerevisiae uronate dehydrogenase.
 XX
 KW uronate reductase; uronate dehydrogenase; Saccharomyces cerevisiae;
 KW synthesis; L-ascorbic acid.
 XX
 OS Saccharomyces cerevisiae.
 XX
 PN DE19604798-A1.
 XX
 PD 14-AUG-1997.
 XX
 PF 09-FEB-1996; 96DE-01004798.
 XX
 PR 09-FEB-1996; 96DE-01004798.
 XX
 PA (HERB-) HERBSTREITH & FOX PEKTIN-FAB NEUENBUERG.
 XX

PI Mattes R, Kulbe K;
 XX
 DR WPI; 1997-403701/38.
 DR N-PSDB; AAT87004.
 XX
 PT Recombinant uronate reductase and L-gulono-gamma-lactone oxidase - used
 PT in L-ascorbic acid production from D-glucuronic or galacturonic acids.
 XX
 XX Claim 27; Page 18-19; 26pp; German.
 XX
 CC The present sequence represents a uronate dehydrogenase (reductase)
 CC isolated from *Saccharomyces cerevisiae*, which is used for synthesis of L-
 CC ascorbic acid. DNA encoding L-gulono-gamma-lactone oxidase (GluOx) is
 CC also used in the production of L-ascorbic acid. The uronate dehydrogenase
 CC has higher affinity for substrate (D-glucuronic or galacturonic acids)
 CC and coenzymes than the similar enzyme from *Lipomyces starkeyi*, also
 CC better stability and greater yield when produced recombinantly
 XX
 SQ Sequence 312 AA;

Query Match 33.2%; Score 574.5; DB 2; Length 312;
 Best Local Similarity 40.8%; Pred. No. 1.9e-47;
 Matches 127; Conservative 54; Mismatches 85; Indels 45; Gaps 7;
 QY 5 KTFTLSNGVKIPGVGFTFASGSKGTYTAVTTALKTYRHLDCAWYILNEGEVGEIGIR 64
 Db 10 KILSLNTGAQIPQICGLGTWQSK--ENDAYKAVLTALKDGYRHIDTAIYRNEDQVQAIAK 67
 QY 65 DFLKENSVPKREDIFVCTKWNHLHRYEDVLSIDSLKRLGLDYDMFLVHWPAAEKN 124
 Db 68 D-----SGVPREEIFVTIKLWCTQHHEPEV--ALDQSLKRLGLDYDLYLMHWPAAEL--- 117
 QY 125 GQGEPKIGPDKGVILKDLTENPE-----PTWRAMEKIYEDRKARSIGV 168
 Db 118 -----DPAYIKNEDILSVPTKDGSRADVITNNFNKTMQLMQLPCKTKTAVGV 168
 QY 169 SNWTIADLEKM--SKFAKVMPHANQIEIHPPLNEELVOYCFCKNIMPVAYSPLGSONQV 226
 Db 169 SNFSINNKLKOLLASQGNKLTAAQVVEIHPLLPQDELINFCCKSGIWEAYSPLGS---- 224
 QY 227 PTTGERSVENKTLNIEAEKGNNTLAQVLIANGLRGYYVLPKSNPKRIESNFKSIELSD 286
 Db 225 --TDAPLKEPVEILEIAKNNVQPGHVVISWHVQRGYVLPKSNPNDRIKTRKIFTLTST 282
 QY 287 ADPEAINAVAK 297
 Db 283 EDFAINNISK 293

RESULT 5
 AAG63565
 ID AAG63565 standard; protein; 312 AA.
 XX
 AC AAG63565;
 XX
 DT 15-OCT-2001 (first entry)
 XX
 DE Amino acid sequence of a nuclear gene encoding a yeast protein.
 XX
 KW Methionine gamma-lyase; mdeA gene; free folding energy; gene shuffling;
 KW directed evolution; molecular breeding.
 XX
 OS Saccharomyces cerevisiae.
 XX
 PN WO200155342-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 31-JAN-2001; 2001WO-US003186.
 XX
 PR 31-JAN-2000; 2000US-00494921.
 PR 08-DEC-2000; 2000US-00734237.
 XX

Query Match 100.0%; Score 1731; DB 5; Length 325;
 Best Local Similarity 100.0%; Pred. No. 5.5e-161;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNGKTFTLSNGVKIPGVGFTFASGSKGETYTAVTALTGTGVRHLDCAWYILNEGEVG 60
 DB 1 MSNGKTFTLSNGVKIPGVGFTFASGSKGETYTAVTALTGTGVRHLDCAWYILNEGEVG 60

QY 61 EGIRDFLKENPSVKREDIFVCTKWNHLHRYEDVLSIDDSLKRLGLDYDMFLVHWP 120
 DB 61 EGIRDFLKENPSVKREDIFVCTKWNHLHRYEDVLSIDDSLKRLGLDYDMFLVHWP 120

QY 121 AEKNGQGEKPGCKYVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMS 180
 DB 121 AEKNGQGEKPGCKYVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMS 180

QY 181 KFAKVMPHANQIEIHPFLPNEELVQYCFSKNIMPVAYSPGSGNQVPTTGERVSENKTLN 240
 DB 181 KFAKVMPHANQIEIHPFLPNEELVQYCFSKNIMPVAYSPGSGNQVPTTGERVSENKTLN 240

QY 241 EIAEKGGNTLAQVLIANGRLRGYVVLPKSSNPKRIESNFKSIELSDADFEAINAVAKGRH 300
 DB 241 EIAEKGGNTLAQVLIANGRLRGYVVLPKSSNPKRIESNFKSIELSDADFEAINAVAKGRH 300

QY 301 FRFVNKMDTFGYDVPETAKNLSA 325
 DB 301 FRFVNKMDTFGYDVPETAKNLSA 325

RESULT 2

ADE39629
 ID ADE39629 standard; protein; 325 AA.
 AC ADE39629;
 DT 29-JAN-2004 (first entry)
 DE Penicillium reducing enzyme protein, SEQ ID No 3.
 KW optically active 2-hydroxycycloalkane carboxylic acid ester;
 KM 2-oxocycloalkane carboxylic acid ester; enzyme; Penicillium; reduction.
 OS Penicillium citrinum.
 XX EPI323827-A2.
 XX 02-JUL-2003.
 XX 20-DEC-2002; 2002EP-00258814.
 XX 27-DEC-2001; 2001JP-00395884.
 XX 27-DEC-2001; 2001JP-00395885.
 XX 10-APR-2002; 2002JP-00107648.
 XX (SUMO) SUMITOMO CHEM CO LTD.
 XX Asako H, Wakita R, Itoh N;
 XX WPI; 2003-723302/69.
 XX N-PSDB; ADE39630.

Producing optically active 2-hydroxycycloalkane carboxylic acid ester by reacting 2-oxocycloalkane carboxylic acid ester with transformant having ability to asymmetrically reduce the ester to optically active ester.

Claim 9; SEQ ID NO 3; 45pp; English.

The invention relates to a novel method for producing an optically active 2-hydroxycycloalkane carboxylic acid ester. The novel method involves: allowing 2-oxocycloalkane carboxylic acid ester to react with a transformant, or a dead cell or their extract, artificially provided with the ability to asymmetrically reduce a 2-oxocycloalkane carboxylic acid

CC ester to the optically active 2-hydroxycycloalkane carboxylic acid; and an
 CC ability to regenerate a coenzyme on which an enzyme having the above
 CC ability depends; and collecting the resulting optically active 2-
 CC hydroxycycloalkane carboxylic acid. The optically active 2-
 CC hydroxycycloalkane carboxylic acid is useful as an intermediate for the
 CC production of bioactive substances. This sequence represents a
 CC Penicillium enzyme protein used to reduce 2-oxocycloalkane carboxylic acid
 CC ester to the optically active 2-hydroxycycloalkane carboxylic acid as part
 CC of the method of the invention.

XX Sequence 325 AA;

Query Match 100.0%; Score 1731; DB 7; Length 325;
 Best Local Similarity 100.0%; Pred. No. 5.5e-161;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNGKTFTLSNGVKIPGVGFTFASGSKGETYTAVTALTGTGVRHLDCAWYILNEGEVG 60
 DB 1 MSNGKTFTLSNGVKIPGVGFTFASGSKGETYTAVTALTGTGVRHLDCAWYILNEGEVG 60

QY 61 EGIRDFLKENPSVKREDIFVCTKWNHLHRYEDVLSIDDSLKRLGLDYDMFLVHWP 120
 DB 61 EGIRDFLKENPSVKREDIFVCTKWNHLHRYEDVLSIDDSLKRLGLDYDMFLVHWP 120

QY 121 AEKNGQGEKPGCKYVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMS 180
 DB 121 AEKNGQGEKPGCKYVILKDLTENPEPTWRAMEKIYEDRKARSIGVSNWTIADLEKMS 180

QY 181 KFAKVMPHANQIEIHPFLPNEELVQYCFSKNIMPVAYSPGSGNQVPTTGERVSENKTLN 240
 DB 181 KFAKVMPHANQIEIHPFLPNEELVQYCFSKNIMPVAYSPGSGNQVPTTGERVSENKTLN 240

QY 241 EIAEKGGNTLAQVLIANGRLRGYVVLPKSSNPKRIESNFKSIELSDADFEAINAVAKGRH 300
 DB 241 EIAEKGGNTLAQVLIANGRLRGYVVLPKSSNPKRIESNFKSIELSDADFEAINAVAKGRH 300

QY 301 FRFVNKMDTFGYDVPETAKNLSA 325
 DB 301 FRFVNKMDTFGYDVPETAKNLSA 325

RESULT 3

AAW29220
 ID AAW29220 standard; protein; 312 AA.

AC AAW29220;
 DT 02-MAR-1998 (first entry)

DE S. cerevisiae uronate dehydrogenase variant Q29G.

DE uronate reductase; uronate dehydrogenase; Saccharomyces cerevisiae;
 KW synthesis; L-ascorbic acid; variant.

OS Saccharomyces cerevisiae.

XX Key Location/Qualifiers

FT Misc-difference 29

FT /label= Q29G

FT /note= "wild type glutamine replaced by glycine"

XX DE19604798-A1.

XX 14-AUG-1997.

XX 09-FEB-1996; 96DE-01004798.

XX 09-FEB-1996; 96DE-01004798.

XX (HERB-) HERBSTREITH & FOX PEKTIN-FAB NEUENBUERG.

XX Mattes R, Kulbe K;

XX PI

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2004, 13:58:21 ; Search time 60 Seconds
(without alignments)
1530.466 Million cell updates/sec

Title: US-10-004-115B-1

Perfect score: 1731

Sequence: 1 MNGKFTLNSGVKIPGVCF.....MKDTGYDVPETAKNLISA 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1731	100.0	325	5	Abp77965 Protein w
2	1731	100.0	325	7	Ade39629 Penicilli
3	575.5	33.2	312	2	Aaw29220 S. cerevi
4	574.5	33.2	312	2	Aaw29217 S. cerevi
5	574.5	33.2	312	4	Aag63565 Amino aci
6	574.5	33.2	312	5	Abg93198 S. cerevi
7	574.5	33.2	313	4	Aag63566 Synthetic
8	573.5	33.1	312	2	Aaw29218 S. cerevi
9	565.5	32.7	312	2	Aaw29219 S. cerevi
10	559.5	32.3	323	3	Aar96294 Carboxyl
11	555	32.1	313	4	Aag63564 Amino aci
12	550	31.8	295	5	Abg93303 C. albica
13	547.5	31.6	344	5	Abg93050 S. cerevi
14	547.5	31.6	344	5	Aau76345 Yeast D-a
15	547	31.6	328	5	Abp53552 T. megach
16	545.5	31.5	328	5	Abp53551 T. megach
17	543.5	31.4	313	4	Aab47466 G. max al
18	543.5	31.4	313	7	Aae39522 Soybean a
19	541	31.3	322	4	Aag63561 Amino aci
20	538	31.1	330	3	Abp53550 T. megach
21	533.5	30.8	309	3	Aag07671 Arabidops
22	529.5	30.6	309	3	Aag33086 Arabidops
23	529.5	30.6	309	3	Aag50693 Arabidops
24	529	30.6	290	3	Aag31845 Arabidops
25	529	30.6	315	3	Aag40418 Arabidops

26	528	30.5	280	5	Abp30416 Streptoco
27	528	30.5	301	5	Abp27439 Streptoco
28	527.5	30.5	302	4	Abb62062 Drosophil
29	527.5	30.5	309	3	Aag43890 Arabidops
30	525	30.3	279	6	Abu18432 Protein e
31	523.5	30.2	313	2	Aay06231 Alfalfa a
32	522.5	30.2	311	3	Aag17752 Arabidops
33	522.5	30.2	311	3	Aag40415 Arabidops
34	522.5	30.2	311	3	Aag04944 Arabidops
35	521	30.1	300	6	Abu24848 Protein e
36	519.5	30.0	311	3	Aag44798 Zea mays
37	519.5	30.0	364	3	Aag44797 Zea mays
38	518.5	30.0	308	4	Aab47468 Z. mays N
39	518.5	30.0	308	7	Aae39524 Corn NADP
40	518.5	30.0	309	3	Aag09538 Arabidops
41	516.5	29.8	320	4	Abb65274 Drosophil
42	516	29.8	279	6	Abu17544 Protein e
43	516	29.8	280	6	Abu01927 S. pneumo
44	516	29.8	280	6	Abu46138 Protein e
45	516	29.8	281	3	Aay81549 Streptoco

ALIGNMENTS

RESULT 1

ABP77965

ID ABB77965 standard; protein; 325 AA.

XX

AC ABB77965;

XX

DT 22-OCT-2002 (first entry)

XX

DE Protein which is capable of producing (S)-4-bromo-3-hydroxybutanoate.

XX

KW (S)-4-bromo-3-hydroxybutanoate; 4-bromo-3-oxobutanoate; pharmaceuticall;
agrochemical; 4-cyano-3-hydroxybutanoic acid.

XX

OS Penicillium citrinum.

XX

PN EP1213354-A2.

XX

PD 12-JUN-2002.

XX

PF 07-DEC-2001; 2001EP-00310251.

XX

PR 07-DEC-2000; 2000JP-00372704.

PR

PR 15-JAN-2001; 2001JP-00006144.

PR

PR 02-FEB-2001; 2001JP-00026594.

PR

PR 11-JUN-2001; 2001JP-00175175.

XX

(SUMO) SUMITOMO CHEM CO LTD.

PI

Asako H, Matsumura K, Shimizu M, Ito N, Wakita R;

XX

WPI: 2002-550350/59.

DR

N-PSDB; ABL59376, ABL59397, ABL59398.

DR

Claim 14; Page 29-31; 56pp; English.

PS

The present sequence represents a Penicillium citrinum protein which is capable of producing (S)-4-bromo-3-hydroxybutanoate by asymmetrically reducing 4-bromo-3-oxobutanoate. The protein and polynucleotides are useful for producing optically active (S)-4-bromo-3-hydroxybutanoate, which is useful as an intermediate in the production of pharmaceuticals and agrochemicals. The (S)-4-bromo-3-hydroxybutanoate produced can also be used to produce 4-cyano-3-hydroxybutanoic acid

XX

SQ Sequence 325 AA;